MOLECULAR BIOLOGY ANALYSIS OF SCHIZOPHRENIA PUTATIVE SUSCEPTIBILITY GENES

Navarrete, Katherinne

Awarding institution:
King's College London

MOLECULAR BIOLOGY ANALYSIS OF SCHIZOPHRENIA PUTATIVE SUSCEPTIBILITY GENES

THESIS SUBMITTED FOR THE DEGREE OF

DOCTORATE OF PHILOSOPHY

AT KING'S COLLEGE LONDON

MAY, 2014

KATHERINNE NAVARRETE

ABSTRACT

Schizophrenia is a common, complex disease with a substantial genetic component and an estimated heritability of 80%. The molecular pathology underpinning its genetic susceptibility is largely unknown. The recent genome-wide association studies (GWAS) have revealed common and rare genetic polymorphisms robustly associated with schizophrenia, giving a great boost to the field. A major challenge, which is essential for these findings to be translated into novel treatments, is to elucidate the biological mechanisms by which these genetic variants alter gene function and biological networks, the downstream effects on brain function and in turn disease risk. The main aim of this thesis was to characterize the molecular and cellular mechanisms by which genetic variation in the loci containing the transcription factor 4 (*TCF4*) and vaccinia-related kinase 2 (*VRK2*) genes confer susceptibility to schizophrenia. The work presented in this thesis can be divided into three main modules: i) a bioinformatic module used to gather information from public database to complement our current knowledge and highlight areas of interest for further investigation; ii) knockdown of *VRK2* and *TCF4* mRNA in a neuronal progenitor cell line; and iii) assay of relative allelic expression of *VRK2*.

The findings presented in this thesis provide experimental evidence for the biological effect of TCF4 and VRK2 on cell proliferation in human neural progenitor cells *in vitro*, indicating this as a possible mechanism underlying the influence of these genetic variants on the pathology of schizophrenia. Additionally, allelic expression analyses revealed an effect of genotype at rs2312147 on *VRK2* allelic expression in foetal samples suggesting a molecular risk mechanism for schizophrenia operating during early brain development, long before the evident appearance of the disease.

The result presented in this thesis support the idea that schizophrenia is in part a neurodevelopmental disorder and provides new insight for molecular mechanism underlying the pathology of the disease.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Leonard C. Schalkwyk, Professor David A. Collier and Dr. Nicholas Bray for allowing me the opportunity to be their PhD student and for believe in this project. I would like to express my gratitude to Dr. Nick for opening the doors of his lab and eventually adopt me as his student. I am very grateful for all his help and support, for guiding me through my PhD and being a source of inspiration and great motivation. It was an honour working with you.

I would like to thank Dr Matt Hill, Dr. Aleks Maruszak for their support and advice in my experimental work. Also, thank you to my colleagues at the CCBB for making my days in the lab more enjoyable ☺.

Thanks to all my friends at the SGDP department for those lovely and long lunch conversations and our unforgettable “Friday cake”.

I would like to give special thank you to my mother and sisters for their love and support.

Finally, I would like to thank my family, my little girl Zazil for being my sunshine and infinity source of kisses and love ☺, and to the little one that is coming, for accompany me in those long nights of writing up and motivate me with his kicks ☺. And of course, a massive thank you to my husband Inti, without you this dream would have not been possible, thank you for all your love and support for being a source of inspiration and motivation and for believe in me.

Table of Contents

Table of Figures

Table of Tables

Chapter 1 General Introduction

1.1 Introduction

Schizophrenia is a complex disease with a strong genetic component. Recent genome-wide genetic association (GWA) studies have revealed a number of discrete genomic regions harbouring common, low risk schizophrenia susceptibility alleles. However, these susceptibility alleles so far only explain a relatively small proportion of the population risk and its genetic component, although this is expected to change as samples sizes and power increase. A number of rare susceptibility variants of moderate effect, including copy number variants and point mutations, have also been identified. In order to translate these genetic findings into clinical practice, disease susceptibility alleles have to be understood in terms of the biological mechanisms mediating their risk. This is currently the greatest challenge faced after identifying a disease locus. Although a great effort have been made in trying to understand the epidemiology and neuropsychology of schizophrenia, there is still little knowledge about the molecular pathology underlying the disease.

My project started from two loci with compelling statistical evidence of association with schizophrenia from GWA studies and I sought to use molecular and cellular biology to further our biological understanding of the associated genetic variants and their putative effector genes. The project focused on loci containing transcription factor 4 (*TCF4*) and vaccinia-related kinase 2 (*VRK2*) genes. Here I describe two pieces of evidence. Firstly, the *cis*-regulatory effect of the risk allele on the expression of *VRK2*, and, secondly, the cellular phenotype produced upon knockdown of *VRK2* or *TCF4* on neuronal progenitor cells. My results show that both genes are involved in regulating neuronal progenitor cell proliferation and that *VRK2* has complex *cis*-regulatory effects involving genetic variants in addition to the risk allele. These results, especially those related to cell proliferation, bear important implications to our understanding of schizophrenia, which I discuss throughout my thesis.

1.2 Schizophrenia

Schizophrenia (MIM 181500) is a severe, long-term and common psychiatric disorder which is ranked in the top 10 causes of disability worldwide (Whiteford et al., 2013). The onset of symptoms typically occurs in early adulthood before the age of 25 (Lewis and Lieberman, 2000). The incidence rate of schizophrenia varies (Dixon et al., 2007a) between 7.7 and 43.0 per 100,000 (McGrath et al., 2008), incidence in England from 1950-2009 was 15 people per 100 000 per year (Kirkbride et al., 2012), with a lifetime prevalence of schizophrenia estimated to be 0.4% (Saha et al., 2005).

Schizophrenia imposes a large financial cost to the health system and to society, due to its early onset, severity and chronic nature. Individuals with schizophrenia have a life expectancy 12 to 15 years shorter as a result of higher suicide rate and ill physical health (Crump et al., 2013; van Os and Kapur, 2009). Common comorbidity with other psychiatric disorders includes depression, anxiety and substance abuse, the latter with a lifetime occurrence of 50% (Buckley et al., 2009).

Due to the absence of physical biomarkers the diagnosis of schizophrenia is based on the presence of behavioural symptoms and is made by reference to the criteria in the fourth edition of the American Psychiatric Association's Diagnostic and Statistical Manual (DSM-IV) (A.P. Association, 2000) and the International Classification of disease (ICD-10).

The disease is characterised by a range of psychological symptoms that are commonly divided into three main categories: i) psychotic or 'positive' symptoms such as auditory hallucinations, delusions, and thought disorder; ii) deficit or 'negative' symptoms including severe disturbances in social interactions and expression of affect, loss of motivation, inability to experience pleasure, poverty of speech and lack of interest and initiative; iii) cognitive dysfunction affecting attention, working memory, and executive and intellectual function (Andreasen, 1995; Tamminga and Holcomb, 2005).

The main treatments for schizophrenia are antipsychotic drugs, which have affinity for a number of neurotransmitter receptors in the brain. The common mechanism of action is blockade of Dopamine D2 receptors (Creese et al., 1976; Seeman et al., 1975), which remain sufficient and necessary for antipsychotic efficacy. Although, antipsychotic are largely effective in treating the positive symptoms, they are unsuccessful for treating negative and cognitive symptoms, which are more persistent. Additionally, antipsychotics drugs have undesirable side effects, such as motor symptoms and weight gain.

A better understanding of the aetiology and molecular pathophysiology of schizophrenia has the potential to enable the development of better treatments that address the underlying neurobiology of schizophrenia.

1.3 The neurobiology of schizophrenia

The aetiology and pathophysiology of schizophrenia remain largely elusive. However there is evidence for neuropathological, neurochemical as well as structural and functional disturbances in the brain of schizophrenic patients.

Several findings from neuroimaging, neuropathology and neurochemistry studies have reproducibly showed association with the disease.

1.3.1 Neuroimaging

Structural neuroimaging studies have shown structural brain abnormalities in schizophrenia patients compared with controls, including enlarged ventricles and an overall reduction by ~3% of the whole brain volume, with more pronounced reductions in the temporal lobe (Shepherd et al., 2012). The presence of these abnormalities at the onset of the disease is consistent with the neurodevelopmental hypothesis of the disease (Weinberger, 1987).

Functional imaging studies have been used to study brain regions that may be dysfunctional in schizophrenic patients. These studies are used to measure the activity of

the brain under different conditions; for example, comparing patients and controls at rest or during cognitive tasks, or correlating brain activity with psychiatric symptoms across patients or within a patient over time (Fu and McGuire, 1999). A reduced activation of the prefrontal cortex and temporal lobe appears to be the most consistent finding in schizophrenia (Ragland et al., 2007). The many studies conducted in this area to date suggest that the observed abnormal function observed in schizophrenic is related to abnormal function in a network of brain regions and might be correlated with structural abnormalities, regional connectivity and symptoms clusters (Niznikiewicz, 2003).

1.3.2 Neuropathology

Neuropathological studies conducted in post-mortem brain from schizophrenic patients have identified subtle cytoarchitectural alterations in the prefrontal cortex and hippocampus (Harrison, 1999). Selemon and colleagues (Selemon and Goldman-Rakic, 1999) described an increased density of pyramidal cells in the prefrontal cortex without neuronal loss. This alteration could be attributable to a reduction in the neuropil (interneural space) or to a decrease in somal volume and in dendritic spines observed in schizophrenic patients (Black et al., 2004; Guidotti et al., 2000; Lewis et al., 2003b; Selemon and Goldman-Rakic, 1999).

Additionally, studies have shown that the numbers of GABAergic cortical interneurons are reduced in schizophrenia (Benes, 1993). This correlates with findings showing reduced levels of mRNA and protein of glutamic acid decarboxylase (GAD) 67; the principal synthesizing enzyme for GABA in the cortex (Akbarian et al., 1995; Guidotti et al., 2000; Volk et al., 2000). The reduction of GAD67 is also associated with decreased levels of mRNA of other genes expressed in GABA neurons such as reelin (Impagnatiello et al., 1998) and parvalbumin (Beasley and Reynolds, 1997). Alteration in the GABA system might explain the abnormal activity observed in the frontal cortex in neuroimaging studies of schizophrenia patients as well as the neuronal connectivity impairment associated with schizophrenia.

Additional neuropathological studies have shown an abnormal expression of genes associated with process of myelination (Hakak et al., 2001) and a reduction in the number of oligodendrocytes in schizophrenia (Byne et al., 2006; Vostrikov et al., 2007). Neuropathological findings have promoted the view of schizophrenia as a disorder of connectivity and of the synapse as well as supporting the neurodevelopmental hypothesis of schizophrenia (see section 1.4)

1.3.3 Neurochemistry

The neurochemical hypotheses of schizophrenia spawned from observations that some psychotropic drugs, such as amphetamine, phencyclidine or lysergic acid diethylamide (LSD), have the ability to induce or mimic in normal individuals the psychotic symptoms observed in schizophrenic patients, while antipsychotic drugs ameliorate these symptoms. These observations lead to the investigation of altered neurotransmitter systems in schizophrenia.

The dopamine hypothesis postulates that psychotic symptoms of schizophrenia are due to a hyperactivity of the dopaminergic system (Howes and Kapur, 2009). This hypothesis originated from observations that antipsychotic drugs are dopamine (D2) receptors antagonists (Creese et al., 1976; Seeman et al., 1975) and that psychotic symptoms can be generated by dopamine releasing agents such as amphetamines (P.H., 1958). Evidence from positron-emission tomography (PET) studies have shown an elevated density of D2 receptors in the striatum in schizophrenic patients as well as elevated DOPA decarboxylase activity (Joyce and Meador-Woodruff, 1997; Laruelle, 1998). These studies have also shown elevated sub-cortical dopamine release in schizophrenia and confirmed the blockade of D2 receptors by antipsychotics drugs (Breier et al., 1997; Frankle and Laruelle, 2002; Laruelle, 1998; Laruelle et al., 1996).

The glutamate hypothesis was driven by findings showing antagonists of the NMDA (N-methyl-D-aspartate) receptor, such as phencyclidine and ketamine, are capable of inducing psychotic-like symptoms similar to those observed in schizophrenic patients

(Halberstadt, 1995; Javitt and Zukin, 1991; Lahti et al., 1995; Olney et al., 1999). Glutamatergic anomalies such as altered expression of glutamate receptor subunits and binding sites have been detected in post-mortem brain of schizophrenics patients (Beneyto et al., 2007; Eastwood and Harrison, 2010; Oni-Orisan et al., 2008) suggesting a role for glutamatergic pathways in schizophrenia pathology.

There is evidence for abnormalities involving a wide variety of additional neurochemical systems, for example GABA, serotonin or norepinephrine. Those together with the above described form complex circuits that converge in neurochemical pathways involved the neuropathology of schizophrenia (Carlsson et al., 2001; Lieberman and Koreen, 1993; Tost and Meyer-Lindenberg, 2011).

1.4 The neurodevelopmental hypothesis of schizophrenia

The neurodevelopmental hypothesis of schizophrenia postulates that this disorder owes its origins to pathological processes happening early in life before the onset of the disease (Weinberger, 1987). Neurodevelopmental hypotheses centre on two potential susceptibility periods: the prenatal period and adolescence/early adulthood. In both periods, important developmental brain processes and maturation of areas functionally associated with the disease are occurring. The neurodevelopmental hypothesis is supported by several pieces of evidence provided by neuropathological, epidemiological and clinical studies (Fatemi and Folsom, 2009). As stated in section 1.3.1, the presence of neuroanatomical deviations at the onset of the disorder is considered to be a key piece of evidence in favour of the neurodevelopmental hypothesis. In addition, cytoarchitectural features such as reduced synaptic markers as well as the lack of classic neurodegenerative changes in the brains of patients with schizophrenia are consistent with a neurodevelopmental origin to the disorder (Harrison, 1997). Epidemiological studies have provided indirect evidence for association between obstetric complication, such as gestational diabetes, preeclampsia, abnormal fetal growth and development and complications of delivery, with an increased risk of schizophrenia (Clarke et al., 2006).

Furthermore, maternal malnutrition and infection during pregnancy is also associated with an increased risk of schizophrenia in offspring (Brown and Patterson, 2011; Brown et al., 1996). Other studies have found that adults who become schizophrenic have a high incidence of deficits in cognition, motor skill, attention, and verbal memory as children (Chua and Murray, 1996; Erlenmeyer-Kimling et al., 2000; Walker et al., 1996). Schizophrenic patients also show an increased occurrence of minor physical and cerebral anomalies, suggesting disturbed intrauterine developments (Lawrie, 2001). Recently, neurons derived from induced pluripotent stem cells (iPSCs) of patients with schizophrenia have been shown to display altered development of synaptic connections (Brennand et al., 2011).

1.5 Genetics of schizophrenia

Genetic epidemiology studies have consistently shown a strong genetic component with an estimated heritability around 80% (Cardno et al., 2002; van Os and Kapur, 2009). Family studies have found a higher risk for schizophrenia in first-degree relatives of individuals with the disorder (3-15%) than in the general population (0.5-1%) (Shih et al., 2004). The concordance rate for schizophrenia in monozygotic twins is estimated at 45-75% and 4-15% for dizygotic twins (Lichtermann et al., 2000). However, our knowledge of the underlying aetiology and the precise genetic architecture of the disorder, which gives rise to these high heritability estimates, remain very incomplete. Schizophrenia is thought to be a complex disorder which involves multiple genetic variants, including common polymorphisms of small effect-size and rare mutations of moderate to large effects (Figure 1.1), environmental and epigenetics factors (Cannon et al., 2002; McGrath et al., 2010; Pidsley and Mill, 2011; Rujescu and Collier, 2009; St Clair et al., 2005).

The first attempts to identify risk alleles for schizophrenia were through linkage analyses in multiple affected families or affected sibling pairs. Linkage studies are powerful tools to detect loci which have high penetrance, especially those following Mendelian inheritance, and were very successful in identifying genes for diseases such as Huntington's and

cystic fibrosis. However, they have low statistical power to detect genetic variants of small effect-sizes, which have very low penetrance and are therefore often transmitted between unaffected individuals (Risch and Merikangas, 1996). The second parallel approach involved candidate gene association studies. They focused on analysing genetic variation within and around genes proposed as good candidates based on their known biological function. This study design is better in identifying common variants with small to large effect sizes (depending on sample size used). Both strategies produced many putative disease loci but few had robust evidence for replication, suggesting that large effect variants exists but are largely private to individual families, presence of allelic heterogeneity (multiple risk alleles on same gene) and the absence of common variants of large effect (leading to false positives in small sample size studies).

Due to technological advancements in the field of genetics (the systematic identification of genome-wide genetic variation) and the development of new microarray technologies allowing large scale, low cost genotyping, systematic genome-wide association studies (GWAS) of common genetic variation in large cohorts became possible. GWAS has shed more light on the role of common variation in the genetic architecture of schizophrenia and have provided robust evidence for association with multiple loci. It has consistently shown that common variants conferring risk to schizophrenia do not have large effects and that large sample sizes, in the order of tens or hundred of thousands of individuals, are needed to identify risk loci robustly. Genetic studies in families or large population cohorts have also shown that rare variation can have large effects and represents a major part of genetic susceptibility to schizophrenia. Currently, large scale studies based on DNA sequencing studies are being performed to better quantify the effects of rare variation.

I will briefly review the main findings from linkage, candidate genes and GWAS. This does not aim to be a comprehensive review but to summarise the most important genetics findings to date in schizophrenia.

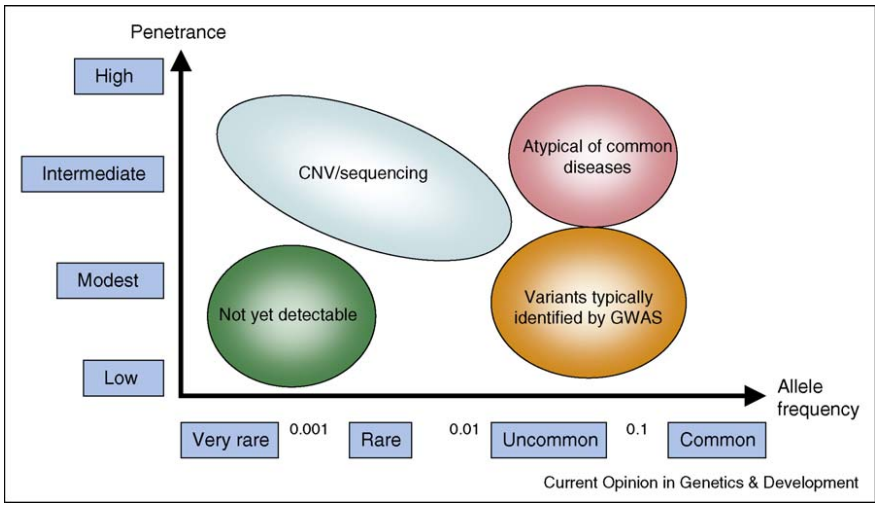


Figure 1.1 Schematic representation of current genetic findings in schizophrenia.

The frequency and penetrance of risk alleles determine which methodology can be used to detect them. Common variants with small effect size associated with schizophrenia have been identified through GWAS studies, while uncommon and rare mutations of large effect such as CNVs, can be identified by sequencing (taken from Owen et al., (Owen et al., 2009)).

1.5.1 Genetics studies

1.5.1.1 Linkage

In linkage studies, families with at least two affected individuals are analysed to identify the co-segregation of a chromosomal region with the disease phenotype (Lander and Schork, 1994). The linkage study design provides statistical power to detect risk alleles of large effect, usually associated with Mendelian disorders or Mendelian forms of complex traits (Lander and Schork, 1994) (Figure 1.1). Linkage regions can be many megabases in sizes and contain dozens of genes, making difficult the identification of the gene or genes responsible for the phenotype (Altshuler et al., 2008; Lander and Schork, 1994). Mapping is dependent on the identification of recombination events within linkage regions, which can narrow the candidate region to a smaller number of genes which can then be sequenced to identify causative mutations.

Linkage results in schizophrenia have shown poor replication success. To date, over 20 genome-wide linkage studies have been completed, implicating a large number of putative schizophrenia susceptibility loci. Three meta-analyses have been performed showing evidence for linkage on chromosomes 1q, 2q, 3p, 4q, 5q, 6p, 8p, 11q, 13q, 14p, 20q and 22q (Badner and Gershon, 2002; Lewis et al., 2003a; Ng et al., 2009). Notably between the three studies, there is only agreement for a single locus on chromosome 8p. As noted above, linkage analysis identifying broad regions, of millions of centimorgans, but unlike Mendelian disorders, fully penetrant disease variants have not been found for schizophrenia and thus recombination events cannot be used to fine map loci as susceptibility variants for complex disorders have low penetrance. Stefansson et al. attempted to map the linkage signal for schizophrenia on chromosome 8p using linkage disequilibrium – this led the identification of a schizophrenia-associated haplotype within the *neuregulin 1* gene (*NRG1*) (Stefansson et al., 2002) which remains controversial.

One of the strongest evidence for schizophrenia reported by linkage studies came from a family study reporting a reciprocal translocation between 1q42 and 11q14.3 that co-segregates with schizophrenia, bipolar disorder and major depression (St Clair et al., 1990). This translocation affects *disrupted in schizophrenia 1 and 2* (*DISC1 and DISC2*). *DISC1* is to date one of the most studied risk genes for schizophrenia. It is associated with cellular process such as cell division and signal transduction and has been implicated in brain development by functional imaging studies (Brandon and Sawa, 2011).

1.5.1.2 Candidate gene association studies

Association studies assess differences in allele frequencies between affected and non-affected individuals, organized as cases and controls for dichotomous traits (McGuffin et al., 2003). One advantage of this strategy is greater power to detect alleles of smaller effect-size than linkage studies (Risch and Merikangas, 1996) (Figure 1.1). Candidate genes association studies are carried out on genes within chromosomal regions

implicated through linkage studies (positional candidates) (Collins, 1995) or plausible genes based on their known function (functional candidates) (McGuffin et al., 2003). Candidate gene studies are limited due to our poor understanding of the pathophysiology of schizophrenia, and have led to a high number of false positives due to population stratification, multiple testing and the small sample sizes used (Altshuler et al., 2008). Hundreds of genes have been tested for association with schizophrenia (see www.schizophreniaforum.org/res/sczgene for an updated list), the most cited candidates include *DRD2*, *DTNBP1*, *NRG1*, *DISC1*, and *COMT* (Muir et al., 2008; Pawel et al., 2010; Walker et al., 2010; Wirgenes et al., 2010). Currently candidate gene studies are less common and their evidence considered weaker than that of GWAS unless they are done on large sample sizes. Most popular candidate genes do not show evidence of association in large GWAS, suggesting that the candidate genes approach is not necessarily an effective way to identify schizophrenia susceptibility loci.

1.5.1.3 GWAS for common variation

Genome-wide association studies compare the frequencies of single nucleotide polymorphisms (SNPs) spanning the whole genome between cases and unaffected control individuals for a disease to identify loci associated with disease susceptibility (Corvin et al., 2010). GWAS provides a powerful tool to identify alleles associated with complex disorders. Hundreds of associations with many traits have been identified using GWAS (see a compendium at NIHGRI GWAS Catalogue, www.genome.gov/gwastudies). A review of the NIHGRI GWAS Catalogue (visited on 20th of April 2014) shows 31 GWAS for schizophrenia, referring to more than 60 SNPs suspected to be involved in the pathogenesis of schizophrenia at an association p-value of $<5 \times 10^{-8}$ (as per Dudbridge and Gusnanto, (Dudbridge and Gusnanto, 2008)), and pointing to more than 100 genes. However, the biological mechanism linking these genetic variations to the pathology of schizophrenia is still unknown. Table 1.1 summarize GWAS study finding for schizophrenia and genes/loci significantly associated with risk at *P* values $< 5 \times 10^{-8}$.

Several genomic loci showing association contain plausible candidates for schizophrenia pathology, for example, *ZNF804A*, *TCF4*, *MIR137*, *CACNA1C*, *NRGN* and the *MHC* region, which includes a large region of strong linkage disequilibrium (LD) with signals from HLA, histone genes and possibly others non-immune system genes such as *NOTCH4* (International Schizophrenia et al., 2009; Stefansson et al., 2009). Biological functions reported for these genes support their plausible role in schizophrenia susceptibility, such as calcium signalling pathways, neuronal plasticity, synaptogenesis and neurodevelopment (Giegling et al., 2010; Melom and Littleton, 2011; Silber et al., 2008; Smrt et al., 2010).

Table 1.1 Schizophrenia genome-wide association studies and genes/loci significantly associated with risk at *P* value < $5x10^{-8}$.

Reference	Region	Reported Gene(s)	Strongest SNP-Risk Allele	*P value*	OR or beta
(Stefansson et al., 2009)	6p21.32	MHC, NOTCH4	rs3131296-G	2E-10	1.19
	6p22.1	MHC, PRSS16	rs6932590-T	1E-12	1.16
	11q24.2	NRGN	rs12807809-T	2E-9	1.15
	18q21.2	TCF4	rs9960767-C	4E-9	1.23
(Shi et al., 2009)	6p21.32	HLA-DQA1	rs9272219-G	7E-8	1.14
	6p22.1	SLC17A1, SLC17A3, BTN3A2, BTN2A2, BTN3A1, HIST1H2AG, HIST1H2BJ, PRSS16, POM121L2, ZNF184	rs13194053-T	1E-8	1.28
(International Schizophrenia et al., 2009)	6p22.1	MHC	rs13194053-T	1E-8	1.22
(Schizophrenia Psychiatric Genome-Wide Association Study, 2011)	18q21.2	CCDC68	rs12966547-G	3E-8	1.40
	10q24.32	CNNM2	rs7914558-G	2E-8	1.22
	8p23.2	CSMD1	rs10503253-A	2E-8	1.16
	3p21.1	ITIH3, ITIH4	rs2239547-?	6E-8	1.10
	1p21.3	MIR137	rs1625579-T	2E-11	1.12

	1q43	NR	rs6703335-?	5E-8	1.09
	10q24.33	NT5C2	rs11191580-T	3E-8	1.20
	2q32.3	PCGEM1	rs17662626-A	5E-8	1.20
	6p22.1	TRIM26	rs2021722-C	2E-12	1.15
(Yue et al., 2011)	6p22.1	NKAPL	rs1635-?	7E-12	1.28
	11p11.2	TSPAN18	rs11038167-?	1E-11	1.29
(Shi et al., 2011)	1q24.2	BRP44, DCAF6	rs10489202-A	1E-8	1.23
	8p11.23	LSM1, WHSC1L1	rs16887244-?	1E-10	1.19
(Bergen et al., 2012)	2q37.2	CENTG2	rs13025591-?	8E-8	1.11
	5q12.1	Intergenic	rs7709645-?	4E-8	1.11
	7p22.3	MAD1L1, SNX8, NUDT1, FTSJ2	rs12666575-?	2E-9	1.12
	6p22.1	MHC	rs17693963-?	3E-11	1.24
	8q21.3	MMP16	rs7004633-?	6E-8	1.15
	10q24.33	NT5C2, CNNM2, other genes	rs11191580-?	2E-9	1.23
(Irish Schizophrenia Genomics and the Wellcome Trust Case Control, 2012)	6p22.1	MHC, TRIM26	rs2523722-G	1E-16	1.25
(Betcheva et al., 2013)	1q32.2	HHAT	rs7527939-C	6E-9	2.63
(Aberg et al., 2013)	18q21.2	TCF4	rs1261117-T	3E-10	1.60
(Ripke et al., 2013)	1q43	AKT3, CEP170	rs14403-C	2E-8	1.10
	10q24.32	C10orf32, AS3MT, CALHM1, CALHM2, CALHM3, CNNM2, CYP17A1, INA, MIR1307, NT5C2, PCGF6, PDCD11, SFXN2, ST13P13, TAF5, USMG5, WBP1L	rs7085104-A	4E-13	1.11
	12q24.31	C12orf65, ABCB9, ARL6IP4, CDK2AP1, MIR4304, MPHOSPH9, OGFOD2, PITPNM2, RILPL2, SBNO1,	rs11532322-A	2E-8	1.09

		SETD8			
	2q37.1	C2orf82, GIGYF2, KCNJ13, NGEF	rs778371-G	2E-8	1.09
	12p13.33	CACNA1C	rs1006737-A	5E-12	1.10
	10p12.31	CACNB2, NSUN6	rs17691888-G	1E-10	1.16
	2q33.1	FONG, C2orf47, C2orf69, SPATS2L, TYW5	rs2949006-T	1E-8	1.10
	5q33.1	GRIA1	rs17504622-T	3E-9	1.24
	6p21.32	HLA-DRB9	rs114002140-A	9E-14	1.67
(Ripke et al., 2013)	1p31.1	Intergenic	rs10789369-A	4E-10	1.10
	5q33.1	Intergenic	rs2910032-C	4E-8	1.08
	2q32.1	Intergenic	rs4380187-A	6E-8	1.08
	17p13.3	Intergenic	rs4523957-T	6E-8	1.08
	3p22.2	Intergenic	rs6550435-G	6E-8	1.08
	3p21.1	ITIH3, ALAS1, ALDOAP1, BAP1, C3orf78, DNAH1, GLT8D1, GLYCTK, GNL3, ITIH1, ITIH4, MIR135A1, MIRLET7G, MUSTN1, NEK4, NISCH, NT5DC2, PBRM1, PHF7, PPM1M, RFT1, SEMA3G, SFMBT1, SPCS1, STAB1, TLR9, TMEM110, TNNC1, TWF2, WDR82	rs4687552-T	1E-8	1.09
	7p22.3	MAD1L1, FTSJ2, NUDT1, SNX8	rs6461049-T	6E-13	1.11
	19p13.11	MAU2, CILP2, GATAD2A, GMIP, HAPLN4, LPAR2, MIR640, NCAN, NDUFA13, PBX4, SUGP1, TM6SF2, TSSK6, YJEFN3	rs2905424-T	3E-9	1.09
	1p21.3	MIR137, DPYD	rs1198588-T	2E-12	1.12
	8q21.3	MMP16	rs11995572-T	3E-8	1.12
	2p22.2	QPCT, C2orf56, CEBPZ, PRKD3, SULT6B1	rs2373000-T	7E-9	1.09
	1q43	SDCCAG8	rs1538774-G	3E-8	1.09
	5q21.1	SLCO6A1	rs6878284-C	9E-9	1.09

	11q25	SNX19	rs7940866-T	2E-9	1.09
	18q21.2	TCF4	rs4801131-C	1E-8	1.08
	8q24.3	TSNARE1	rs4129585-A	2E-10	1.09
	2q22.3	ZEB2	rs12991836-C	1E-8	1.08
	5q12.1	ZSWIM6, C5orf43	rs171748-A	4E-8	1.08
(Wong et al., 2013)	Xq28	RENBP, ARHGAP4, MECP2	rs2269372-A	4E-8	1.31

1.5.1.4 Copy Number Variation

Genomic variants (deletions or duplications of segments of the genome) have traditionally been identified by cytogenetic using chromosome staining and light microscopy. In the more modern terminology, copy number variants (CNVs) usually refer to submicroscopic (i.e., not detected by traditional cytogenetic tools) deletions and duplications of about 500 kb to 1 kb (Conrad et al., 2010; Hastings et al., 2009). CNV constitute a common form of genetic variation and the vast majority of CNVs are inherited (Conrad et al., 2010), however, *de novo* CNVs can also occur, and these have been strongly implicated in the disease. The consequences of CNVs can be a disruption of gene expression and have been associated with a number of traits, especially neurodevelopmental and metabolic disorders (Hurles et al., 2008; Malhotra and Sebat, 2012).

Several CNVs have been associated with susceptibility to schizophrenia; these findings are presented in Table 1.2. The strongest evidence is for a greater burden of rare *de novo* CNVs affecting genes (International Schizophrenia, 2008; Xu et al., 2008). Several studies have shown that these CNVs are enriched for genes with high expression or known role at the synapse and other brain associated cellular functions (for example Stefansson et al., (Stefansson et al., 2008) and Kirov et al., (Kirov et al., 2012)). GWAS technologies can also detect rare copy number variants (CNVs). Stefansson et al. (Stefansson et al., 2008) also showed CNVs are under negative selection, suggesting a significant effect on fitness.

Table 1.2 Summary of CNVs identified in schizophrenia.

Locus	Type	Genes	References
1q21.1	Deletion	11	(Ikeda et al., 2010; International Schizophrenia, 2008; Levinson et al., 2011; Need et al., 2009; Walsh et al., 2008).
1p36.33	Duplication	5	(Rees et al., 2014)
2p16.3	Exonic deletion	*NRXN1*	(Ikeda et al., 2010; Kirov et al., 2009b; Levinson et al., 2011; Magri et al., 2010; Need et al., 2009; Rujescu et al., 2009; Walsh et al., 2008).
3q29	Deletion/duplication	21	(Kirov et al., 2012; Levinson et al., 2011; Magri et al., 2010; Mulle et al., 2010; Vacic et al., 2011).
7q36.3	Exonic deletion	*VIPR2*	(International Schizophrenia, 2008; Levinson et al., 2011).
15q11.2	Deletion	4	(Kirov et al., 2009a; Kirov et al., 2012; Magri et al., 2010; Stefansson et al., 2008).
15q13.3	Deletion	8	(International Schizophrenia, 2008; Kirov et al., 2012; Levinson et al., 2011; Stefansson et al., 2008; Vacic et al., 2011).
16p11.2	Duplication	26	(Glessner et al., 2010; Kirov et al., 2012; Levinson et al., 2011; McCarthy et al., 2009; Vacic et al., 2011; Walsh et al., 2008).
16p12.1	Deletion	7	(Rees et al., 2014)
16p13.1	Duplication	11	(Ikeda et al., 2010; Ingason et al., 2011; Kirov et al., 2009a; Magri et al., 2010).
17q12	Deletion	15	(Kirov et al., 2009a; Magri et al., 2010; Moreno-De-Luca et al., 2010).
22q11.21	Deletion	29-43	(Glessner et al., 2010; International Schizophrenia, 2008; Stefansson et al., 2008; Vacic et al., 2011).

1.5.1.5 Exome sequencing

Recent studies conducted using exome-sequencing technology, have shown that genes previously associated with schizophrenia are enriched for rare variants and *de novo* mutations (Fromer et al., 2014; Purcell et al., 2014). In addition, these studies show that genes with known prominent functions on the synapse and brain are also enriched for rare risk alleles. These gene sets include the voltage-gated calcium ion channels and genes involved in synaptic process, such as NMDAR (N-methyl-D-aspartate receptor), ARC (activity-regulated cytoskeleton-associated) and PSD-95 (postsynaptic density-95) as well as FMRP (fragile X mental retardation protein) targets. These genes sets have been associated with other neurodevelopmental disorders such as autism and intellectual disability and the authors show that they have high postnatal brain expression (Cross-Disorder Group of the Psychiatric Genomics and Genetic Risk Outcome of Psychosis, 2013; Kirov et al., 2012).

These studies provide further evidence for an implication of synaptic disruption as a pathogenic mechanism in schizophrenia. Furthermore, they support the idea of a shared disease mechanism between other neurodevelopmental disorders and schizophrenia and demonstrate a polygenic burden increasing risk for schizophrenia, comprising rare nonsense and de novo mutations distributed across many genes. Interestingly, these studies do not support the idea of rare variation with large effects on schizophrenia susceptibility (Figure 1.1) and suggest that large sample sizes will also be needed to identify association with rare variants.

1.6 Molecular biology studies for schizophrenia GWAS loci

Genetic studies have provided a robust lead to help uncover the biological processes underlying schizophrenia aetiology. The current challenge is to interpret the biological meaning of these findings to inform projects for the development of new treatments for schizophrenia. Molecular studies focusing on genes pointed to by GWAS of schizophrenia have shown that these genes participate in several cellular functions

relevant to the pathology of the disease such as neuronal proliferation, neurogenesis, neuronal differentiation and maturation as well as other cell regulation processes such as circadian cycle (see Table 1.3 present a representative selection of these studies). Furthermore, some of these studies have also shown using animal models that manipulation of these genes impact on behaviour and cognitive functions. This evidence is helping us to delineate the molecular pathways by which genetic variation influencing these genes can have an effect and in turn to highlight which traits they might be modifying. For example, a transgenic mouse moderately overexpressing *TCF4* shows cognitive deficit, a phenotype well described in patients with schizophrenia (Brzózka et al., 2010). This suggests that the molecular pathways in which TCF4 is involved modulate cognitive function in mice, and potentially in humans (human knock-out of TCF4 have profound cognitive phenotype, see section 1.7). Therefore, identifying these molecular pathways and their effect on neuronal and brain systems can point to potential targets to develop therapies oriented to improve or avoid cognitive deficit.

The present thesis focuses on two putative effector genes for schizophrenia - *TCF4* and *VRK2* - as there is very little known about their role in the brain, making them attractive for molecular studies. I now present a brief description of these two genes. A more detailed characterisation will be presented through the thesis.

Table 1.3 Examples of molecular studies for schizophrenia genes.

Gene	Model	Main finding	Main methodology	Main finding cellular pathway	Reference
TCF4	Transgenic mice overexpressing *Tcf4* postnatally in the forebrain	-Tcf4 transgenic mice present cognitive and sensorimotor deficit. - TCF4 interact with neurogenic bHLH factors NEUROD and NDRF in the adult brain - Circadian deregulation of neuronal bHLH factors in adult hippocampus.	- Behavioural analysis - Reporter gene assays performed in primary cortical neurons and Co-immunoprecipitation assays. - Microarray and qRT-PCR	Circadian cycle	(Brzózka et al., 2010).
ZNF804A	Human neural progenitor cell line CTX0E03	Endogenous *ZNF804A* knockdown has a significant effect on the expression of genes involved in cell adhesion. Suggesting a role for ZNF804A in processes such as neural migration, neurite outgrowth and synapse formation.	- RNAi knowdown - Microarray (Illumina HT-12 v4 BeadChip array) - qRT-PCR - Pathway analysis	Cell adhesion	(Hill and Bray, 2012).
MIR137	Human neural progenitor cell line CTX0E03	Overexpression of miR137 lead to a downregulation of miR137 target Downregulated genes in association with miR137 overexpression are enriched for neuronal differentiation	- Overexpression and inhibition of miR-137 using a mirVanaTM miRNA mimic of the miR-137 precursor mirVanaTM miRNA inhibitor, respectively. - Microarray (Illumina HT-12 v4 BeadChip array) - qRT-PCR - Pathway analysis	Neuronal differentiation	(Hill et al., 2014).
MIR137	Mice and adult mice hippocampal neuroprogenitor cells	miR-137 regulate dendritic morphogenesis, phenotypic maturation, and spine development both in brain and cultured primary neurons. miR-137 targets the	- Relative quantification of mature miRNAs by Taqman miRNA real-time PCR -Overexpression of miR-137 in new-born cells of the adult DG using	Neuronal maturation and dendritic morphogenesis	(Smrt et al., 2010)

		Mind bomb one (Mib1) protein. Mib1 is an ubiquitin ligase important for neurodevelopment.	retrovirus-mediated gene delivery - Overexpression of miR-137 in neuroprogenitor cells using lentivirus - 3'-untranslated region dual luciferase assays of miR-137 target mRNA		
DISC1	Mice and mice neuronal progenitor cells	DISC1 regulates neural progenitor proliferation during embryonic brain development and in the adult dentate gyrus. DISC1 exerts its role in proliferation by modulation of GSK3b/b-catenin signalling pathways. DISC1 knockdown in adult mice elicited hyperactive and depressive behaviours.	- *DISC1* expression knockdown *in vivo* and *in vitro* using shRNA lentivirus. - In utero BrdU labeling - Cell proliferation assays - Luciferase assays *in vitro* binding and kinase assays to test DISC1 inhibition of GSK3β through direct association - Behavioural analysis	Neuronal proliferation	(Mao et al., 2009)

1.7 TCF4

TCF4 (transcription factor 4) encodes a basic helix-turn-helix Ephrussi-box transcription factor located at chromosome 18q21.2. It is widely expressed in human tissues and appears to be involved in multiple biological functions in a variety of cell types (Cisse et al., 2008; Engel and Murre, 2001; Muir et al., 2006; Quong et al., 2002; Sobrado et al., 2009; Wikstrom et al., 2008). Additionally, TCF4 is likely to play a major role in neurodevelopment, especially as its haploinsufficiency has been associated with a rare form of syndromic mental retardation, Pitt-Hopkins syndrome (MIM #610954) (Pitt and Hopkins, 1978). However, there is a paucity of information on its roles in the human CNS. It is known to interact with three proteins involved in neurodevelopment, MATH1 (Flora et al., 2007; Gohlke et al., 2008; Ross et al., 2003), HASH1 (Persson et al., 2000) and NEUROD2 (Ravanpay and Olson, 2008). TCF4's interactions with these transcription factors suggest it is involved in an intricate combinatorial regulatory circuit during CNS development.

Three common variants within or close to *TCF4* have been shown to confer risk to schizophrenia in GWAS studies: rs9960767 (Stefansson et al., 2009), rs4309482 (Steinberg et al., 2011), and rs12966547 (Ripke et al., 2011). Both rs4309482 and rs12966547 lie approximately equidistant from *CCDC68* (coiled-coil domain containing 68, aka cutaneous T-cell lymphoma associated antigen, CTLC SE57-1) a gene with an unknown function, and *TCF4* (in addition to a lincRNA (ENST00000507267; AC091103.1) gene which overlaps with *TCF4*) and it is unclear which gene they affect. However, because mutations in *TCF4* are known to cause intellectual disability (Brockschmidt et al., 2007), it is a plausible hypothesis that all three variants affect *TCF4* in some way.

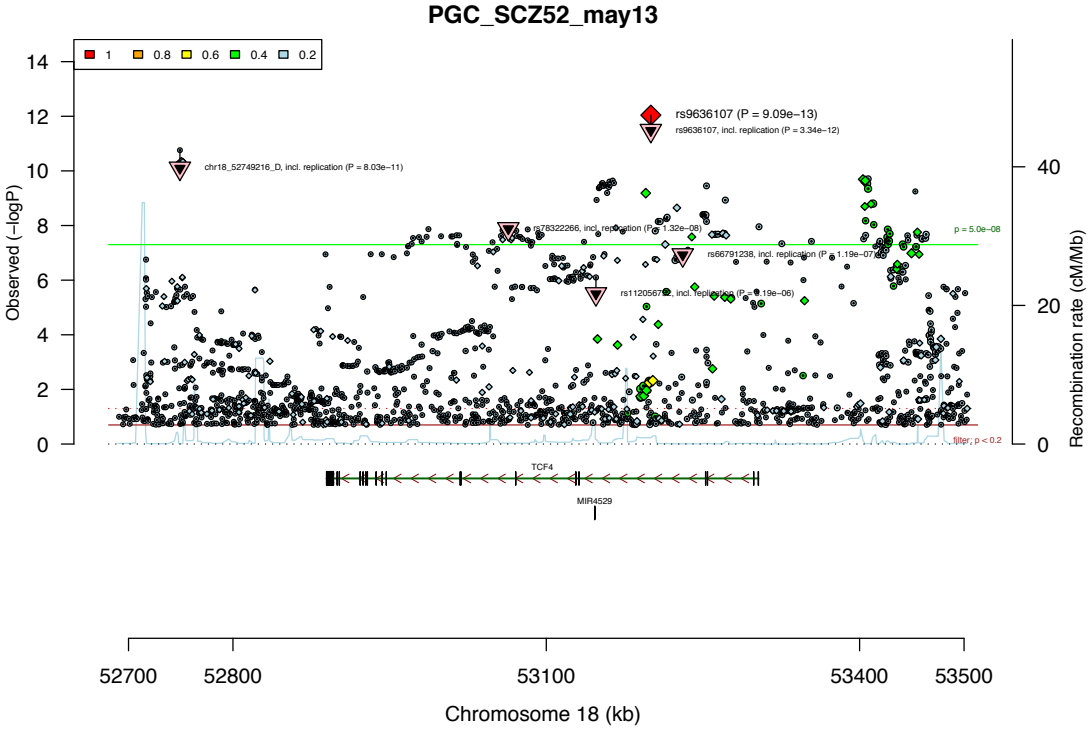


Figure 1.2 Association signals between the *TCF4* markers. Figure obtained from PGC schizophrenia analysis (broadinstitute.org/mpg/ricopilli/)

1.8 VRK2

VRK2 is a member of the vaccinia-related kinase (VRK) family of serine/threonine kinases located at chromosome 2p16.1. Two splice variants generate isoforms A and B of 508 and 397 amino acids, respectively. VRK2A is localised in the cytoplasm with a hydrophobic tail in the C-terminal region that anchors the protein to the membrane of the endoplasmic reticulum and mitochondria (Blanco et al., 2006). VRK2B is shorter and is localised to the nucleus as well as the cytosol (Blanco et al., 2006).

One common variant within *VRK2* has been shown to confer risk to schizophrenia in GWAS (Figure 1.3). The association at rs2312147 was initially reported by Stefansson et al. (Stefansson et al., 2009) with a *P* value not reaching reach genome-wide significance. A follow up study now provides genome-wide evidence of association with rs2312147 (OR = 1.10, $P = 5.4 \times 10^{-10}$) (Steinberg et al., 2011) rs2312147 is located within the first intron of the long isoform of *VRK2* and upstream of the short isoform.

A microdeletion syndrome involving chromosome 2p15-p16.1 has recently been reported in six individuals. They share similar clinical neurodevelopmental and cognitive phenotypes including mental disability and autism as well as facial dysmorphism (Chabchoub et al., 2008; de Leeuw et al., 2008; Felix et al., 2010; Liang et al., 2009; Rajcan-Separovic et al., 2007). Three of the six individuals have microdeletions affecting *VRK2*, but only two presented intellectual disability and autistic features, thus the gene cannot be exclusively responsible for these phenotypes.

Downstream of *VRK2* is *FANCL* (Fanconi anemia, complementation group L), which is known to be involved in Fanconi anemia (MIM ID #227650). Fanconi anemia is a genetically heterogeneous recessive disorder characterised by cytogenetic instability, hypersensitivity to DNA cross-linking agents, increased chromosomal breakage, and defective DNA repair. *FANCL* deficient mice are less fertile and have defective proliferation of germ cells (Zhao et al., 2005). Functionally, *FANCL* is a weaker candidate

for schizophrenia, and the index SNPs in *VRK2* are not in LD with genetic variation within *FANCL*, although it cannot yet be ruled out.

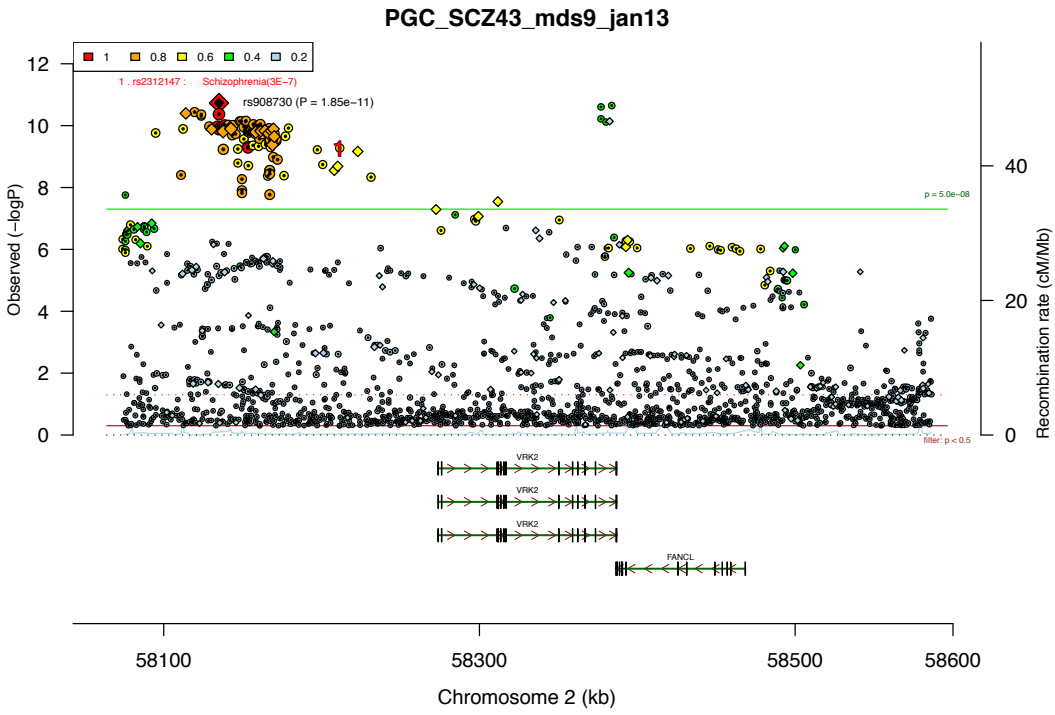


Figure 1.3 Association signals between the *VRK2* markers. Figure obtained from PGC schizophrenia analysis (broadinstitute.org/mpg/ricopilli/).

1.9 Aims of this thesis

This PhD thesis aims to characterise the biological mechanisms through which two schizophrenia susceptibility loci identified through GWAS operate using molecular and cell biology tools. Thus I aim to provide a better understanding of the underlying biology of schizophrenia, especially that related to TCF4 and VRK2, which could point to future avenues for improvements of its diagnosis, management or treatment.

Specific Objectives:

A. *In silico* characterisation of TCF4 through extensive bioinformatic analyses and data mining. This is presented in Chapter 2.
B. Test whether manipulation of the endogenous expression of *TCF4* has an effect on the proliferation of human neural progenitor cells. This is presented in Chapter 3.
C. Test whether *VRK2* expression in the neonatal and adult human brain is influenced by the schizophrenia risk allele (rs2312147 allele G, Steinberg et al., 2011) and other common *cis*-regulatory variation. This is presented in Chapter 4.
D. Assess the effect of *VRK2* knockdown on the transcriptome and proliferation of human neural progenitor cells. This is presented in Chapter 5.

Chapter 2 Bioinformatic analysis of TCF4

The present chapter is a modified version of the work published in: Navarrete K, Pedroso I, De Jong S, Stefansson H, Steinberg S, Stefansson K, Ophoff RA, Schalkwyk LC, Collier DA. 2013. TCF4 (e2-2; ITF2): A schizophrenia-associated gene with pleiotropic effects on human disease. Am J Med Genet Part B 162B: 1–16. The published version of this article and its supplementary material is attached in Appendix 1. I made the major contribution to this published work; I performed all the bioinformatics analysis and contributed to the majority of the writing. The only part of this published work that was not performed by me was the expression analysis of *TCF4* in lymphocytes from schizophrenia patients and controls. This experimental work was performed at Dr Roel Ophoff's Lab by his MSc student Simone de Jong.

2.1 Introduction

The *TCF4* gene on chromosome 18q21.2 (Entrez Gene ID 6925; ensemble ENSG00000196628) encodes a basic helix-turn-helix (bHLH) Ephrussi-box (‘‘E-box’’ or E-) protein transcription factor. The gene spans 437 kb and has 41 exons. It has multiple aliases in addition to the official HUGO name *TCF4* and is also known as *E2-2*, *ITF2*, *PTHS*, *SEF2*, *SEF2-1*, *SEF2-1A*, *SEF2-1B*, *bHLHb19*, *MGC149723*, and *MGC149724*. The pleiotropic disease associations of *TCF4* include rare non-synonymous mutations and gene deletions in Pitt-Hopkins syndrome (PHS), and common risk variants for Fuch’s corneal dystrophy (FCD) and schizophrenia, which are genetically distinct from each other. It should not be confused with *TCF7L2* (transcription factor 7-like 2), another transcription factor mapping to a different locus on chromosome 10q25-25.3, that plays a key role in the Wnt-signaling pathway, which has also been called *TCF4*, and has been associated with schizophrenia (Alkelai et al., 2012; Hansen et al., 2011).

In the present chapter, I performed a systematic literature search for publications describing the TCF4 protein, using the terms E2-2, ITF2, PTHS, SEF2, SEF2-1, SEF2-1A, SEF2-1B, bHLHb19, PHS, and FCD. The literature search was updated up until July 9, 2012, and references collated (see Appendix 1 for full list of references identified for the study). Papers referring to TCF7L2 (but which used the term TCF4) were manually excluded from the list after reading the text to determine this. Several papers whose subject was TCF4 (e2-2) were incorrectly indexed with the term TCF7L2, including the original paper describing the cloning of TCF4 (e2-2). These were retained in the list. Bioinformatic analysis was performed up until April 2012.

2.1.1 The aim of the present study

The aim of this chapter was to characterize TCF4 *in silico* through extensive bioinformatics analyses and data mining and to collect information about TCF4 through a systematic literature review.

2.2 TCF4 and human disease

2.2.1 Pitt-Hopkins Syndrome

Non-synonymous mutations and deletions in the *TCF4* gene are a cause of PHS (MIM #610954), a rare autosomal dominant encephalopathy caused largely by *de novo* mutations (Whalen et al., 2012). PHS was originally described in 1978 in two unrelated patients with mental retardation, recurrent episodes of hyperventilation, and a wide mouth (Pitt and Hopkins, 1978). A series of additional cases helped to refine the clinical features, which are severe psychomotor delay, epilepsy, and daily bouts of diurnal hyperventilation starting in infancy; mild postnatal growth retardation, postnatal microcephaly and distinctive facial features (de Pontual et al., 2009; Orrico et al., 2001; Peippo et al., 2006; Singh, 1993; Van Balkom et al., 2012). Families describe their children with PHS as creative and joyful (pitthopkins.org). Clinical study finds that most have a smiling appearance, as well as anxiety, stereotypic movements, particularly of the arms, wrists, and fingers, and hyperventilation that may be triggered by emotional situations (Whalen et al., 2012). Constipation is common, and there is co-morbidity for the developmental disorder Hirschsprung disease in PHS, a neurocristopathy characterized by the absence of enteric ganglia along a variable length of the intestine.

Using systematic genome-wide BAC-array screening, *de novo* microdeletions on chromosome 18q21.1, which include the *TCF4* gene, were identified in PHS (Amiel et al., 2007; Brockschmidt et al., 2007; Gustavsson et al., 1999; Zweier et al., 2007) and fine mapping of the deleted region led to the identification of heterozygous *TCF4* mutations in non-deleted cases (Amiel et al., 2007; de Pontual et al., 2009), including null alleles created by splice site, frameshift, nonsense mutations, and missense mutations which change highly conserved residues. Almost all mutations are *de novo* and private. Point mutations account for 40% of these, mostly causing premature stop codons, indels 30%, and *de novo* gene deletions 30%. Missense mutations are generally within the bHLH domain, a mutational hotspot. There may also be attenuated loss of function with a less

severe phenotype: one parent carried a mosaic mutation and suffered from depression and epilepsy (de Pontual et al., 2009).

Analysis of the effects of *TCF4* mutations in PHS is consistent with loss of function, with activation of a reporter construct significantly and similarly impaired for nonsense, frameshift, and missense *TCF4* mutants when co-transfected with achaete-scute complex homolog 1 (HASH1; ASCL1) cDNA (de Pontual et al., 2009), a cofactor involved in activation of TCF4 DNA binding (Persson et al., 2000). The causative mutations in *TCF4* (that is, hemizygous deletions and inactivating mutations) indicate that PHS is caused by haploinsufficiency of the gene and that is it is highly dosage-sensitive.

While TCF4 mutations are confirmed as the cause of PHS, autosomal recessive disorders resembling PHS (Pitt-Hopkins- like syndrome) are caused by mutations in NRXN1 and CNTNAP2 (Blake et al., 2010; Zweier et al., 2009). Copy number variants (CNVs) in both of these genes are also associated with schizophrenia and other neurodevelopmental disorders (Friedman et al., 2008; Rujescu and Collier, 2009).

2.2.2 Schizophrenia

Genetic epidemiology studies of schizophrenia have consistently shown a strong genetic component, with estimates of heritability at up to 80% (Cardno et al., 2002; van Os and Kapur, 2009). Multiple genetic variants have been implicated in risk, including common single nucleotide polymorphisms (SNPs) from genome-wide association with small effect-size and rare CNVs of moderate to large effects (Rujescu and Collier, 2009), as well as environmental and epigenetics factors (Pidsley and Mill, 2011). To date, two independent genetic loci in near *TCF4* have been implicated in schizophrenia, one in intron three of the gene and one between the distal end of *TCF4* and the gene *CCDC8*, providing robust GWAS evidence for association (Ripke et al., 2011; Stefansson et al., 2009; Steinberg et al., 2011).

The initial association between TCF4 and schizophrenia was with the SNP rs9960767, in intron three of the gene (P = 4.1 X 10^{-9}; OR 1.23) (Stefansson et al., 2009). A second variant (rs4309482) downstream of the *TCF4* gene reached a genome-wide significance level in a follow-up study (P = 7.8 X 10^{-9}; OR 1.09) (Steinberg et al., 2011), and a third variant (rs12966547) was significantly associated with schizophrenia in a GWAS mega-analysis from the psychiatric GWAS consortium (P = 2.60 X 10^{-10}; OR 1.09) (Ripke et al., 2011). A fourth variant, rs17512836, has also been associated with schizophrenia from the PGC schizophrenia analysis (P = 2.35 X 10^{-8}); after follow-up this gave a combined P value of 1.05 X 10^{-6} (OR 1.23) (Ripke et al., (Ripke et al., 2011); broadinstitute.org/mpg/ricopili/).

The *TCF4* gene, the location of GWAS associated SNPs, and their linkage disequilibrium (LD) with other SNPs at the locus is shown in Figure 2.1. rs12966547 is physically (1.5 kb) very close to rs4309482 and is in complete LD (r^2 = 1), so is effectively the same marker. rs4309482 and rs12966547 are in very weak LD with rs9960767 in the Caucasian population (D' = 0.05, r^2 = 0), that is they are showing independent association signals with schizophrenia. rs17512836 is likewise in very low LD with rs4309482 and rs12966547 (D' = 0.02, r^2 = 0), but in moderate LD with rs9960767 (D' = 1, r^2 = 0.52), indicating the association signal there is not independent.

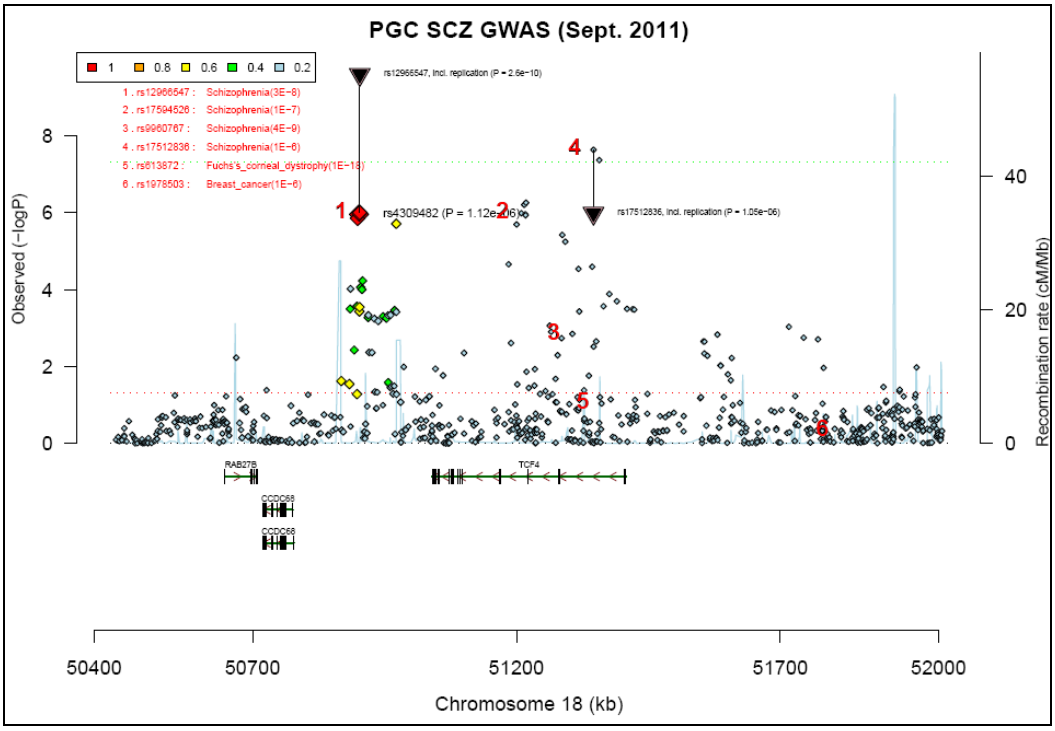


Figure 2.1 Association signals between the TCF4 markers rs4309482, rs12966547, rs9960767, and rs17512836. Figure obtained from PGC schizophrenia analysis (broadinstitute.org/mpg/ricopilli/).

Both rs4309482 and rs12966547 lie approximately equidistant from three genes: *CCDC68* (coiled-coil domain containing 68, aka cutaneous T-cell lymphoma associated antigen, CTLC SE57-1), a gene with unknown function; a long intergenic non-coding RNA (ENST00000507267 or AC091103.1) overlapping *TCF4* in the opposite strand; and *TCF4* itself. Therefore, in contrast with PHS mutations, it is less certain from the GWAS results which gene(s) these variants affect, as they are not obviously functional. However, a parsimonious association with *TCF4* can be hypothesized given that (i) rs9960767 is in an intron of the *TCF4* gene and is in low LD with the other two *TCF4* associated markers, (ii) mutations in *TCF4* are known to cause severe mental illness, and (iii) *TCF4* is target of micro-RNA 137, which has also been associated with schizophrenia (Kwon et al., 2013; Ripke et al., 2011). Therefore, it is reasonable that the gene mediating the genetic

risk conferred by these variants is *TCF4* and that the GWAS-associated variants are tagging two independent *TCF4* loci which have a functional effect on the gene in some way.

However, these SNPs do not show any association with bipolar affective disorder ((Psychiatric, 2011); http://www.broadinstitute.org/mpg/ ricopili/), making them unusual in that many genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross traditional diagnostic boundaries (Williams et al., 2011a). It is also interesting that both common and rare variants affect the same gene and confer susceptibility to brain disorders of varying severity, as it has been shown for other traits (Johansen et al., 2010; Zaghloul and Katsanis, 2010).

Two earlier genetic studies, before the advent of GWAS, examined this locus in schizophrenia: one study examined an unstable CTG repeat (CTG18.1) in an intron of *TCF4* (McInnis et al., 2000) which did not show association, and a second, cytogenetic, study of families with schizophrenia or bipolar disorder found independent pericentric inversions on chromosome 18 close to the TCF4 gene (inv[18][p11.31;q21.2] and inv[18][p11.31;q21.1]) which could conceivably affect gene function (Pickard et al., 2005).

Consistent with its involvement in schizophrenia pathogenesis, blood TCF4 levels were found to be reduced during psychotic states (Kurian et al., 2011), and a small study of post-mortem cerebellar cortices of 14 schizophrenia patients and six controls also highlighted TCF4, but found up-regulation of expression in patients (Mudge et al., 2008). However, post-mortem brain mRNA-level analyses of three adult brain regions, the dorsolateral prefrontal cortex (DLPFC; Brodmann area BA46) (Narayan et al., 2008), the lateral and caudal superior temporal gyrus (BA22) (Barnes et al., 2011) and the frontopolar prefrontal cortex (BA10) (Maycox et al., 2009), found no association between schizophrenia and TCF4. These discrepancies may well be explained by differences in expression patterns among tissues and cell type. Nonetheless, it remains unclear if mRNA expression changes are part of the schizophrenia pathology. In order to gather

additional evidence, colleagues at Dr Roel Ophoff's Lab examined TCF4 relative mRNA levels experimentally in lymphocytes from cases with schizophrenia (n = 106) and controls (n = 96), and found approximately 20% lower expression in patients (fold change = 0.84 and *P*-value = 2.13 10^{-7}, see section 2.9 for details). Williams et al., (Williams et al., 2011b) examined *TCF4 cis*-regulated variation in gene expression for correlation with one of the associated SNPs, rs9960767, using an assay to detect differential allelic expression. They found no evidence for *cis*-regulated mRNA expression in adult cerebral cortex related to genotype at the schizophrenia associated SNP. However, they could not rule out temporal- or tissue-specific effects on expression. Future studies on larger sample sizes looking at specific brain regions across development will be needed to clarify the role of mRNA-level changes of TCF4 in schizophrenia risk.

2.2.3 Fuch's Corneal Dystrophy

Genetic variation within *TCF4* has also been associated with FCD (Baratz et al., 2010; Kuot et al., 2012; Li et al., 2011), a common eye disease affecting about 5% of the population, and the leading cause of corneal transplant (accounting for 10% of all corneal dystrophies) (Lang and Naumann, 1987); 80% of those affected are female. FCD is characterized by the development of collagen-free regions, which may be caused by cell death (Meek et al., 2003). Baratz et al., (Baratz et al., 2010), in a GWAS study, showed that SNPs at the *TCF4* locus were strongly associated with typical FCD, with an odds ratio of 30 for homozygotes for the disease variants. While many markers showed association, the SNPs rs17595731 and rs613872, intronic for *TCF4*, and rs9954153 and rs2286812, upstream of *TCF4*, demonstrated independent signals across the gene and gave the best case–control discrimination. rs613872 showed the strongest evidence for association (*P* =1 X 10^{-18}).

The FCD-associated variants appear partly independent of schizophrenia associations, even though the FCD index SNP rs613872 maps to the same intron of the gene as some of the schizophrenia associations. rs613872 shows weak to moderate LD with rs17512836 (D' = 1, r^2 = 0.25) and rs9960767 (D' = 0.65, 1, r^2 = 0.20) but negligible LD

with rs4309482 (D' = 0.07, r^2 = 0) and rs12966547 (D' = 0.07, r^2 = 0). Given that schizophrenia and FCD are common in the population (about 1%) and may partially share some genetic risk, it will be worth determining whether there is any co-morbidity between these two disorders. However, there is no evidence of this from the literature.

2.3 The *TCF4* gene

TCF4 (immunoglobulin transcription factor 2; *ITF2*; *SEF2*; OMIM *602272) was initially identified in 1990 through cloning and sequencing of the cDNA, identified as encoding a HLH transcription factor binding to Ephrussi, or E box elements (the mu-E5 motif of the immunoglobulin heavy chain enhancer and the kappa-E2 motif of the light chain enhancer) (Henthorn et al., 1990). TCF4 was independently identified by Corneliussen et al. (Corneliussen et al., 1991), who characterized a protein family with homology to other bHLH transcription factors through their binding to glucocorticoid response element E-box motifs in the in the murine leukemia virus SL3-3 enhancer. Corneliussen et al., (Corneliussen et al., 1991) called the new protein family SEF2, for ''SL3-3 enhancer factors''; TCF4 was named SEF2-1B (667 amino acids). They also identified multiple splice variants. Pscherer et al., (Pscherer et al., 1996) identified the murine homologue of TCF4 (SEF2-1B; mouse Sef2) through its binding to the E box of the somatostatin receptor (SSTR2) initiator element, where it can recruit the basal transcription factor TFIIB through physical interaction with SEF2.

2.3.1 Comparative Genomics of TCF4

Comparative genomic analyses on 14 representative vertebrates showed that the length of the TCF4 gene greatly differs even between closely related species, for example medaka and fugu and mouse and rat, probably as a result of the extent of repeats. A tendency towards increase in the size of the gene is detected from fish (36.2 kb) to human (437 kb), which could be explained under neutral evolution (Lynch, 2007). The length of deduced human proteins from alternative *TCF4* transcripts based on the Ensembl annotation is between 511 and 773 amino acids. Although the length of the deduced protein differs across species there is high sequence identity of the primary

sequence, including the DNA binding domain Helix-loop-helix (Pfam domain id: PF00010) (Table 2.1).

Table 2.1 TCF4 gene in 14 species.

Species	Gene ID	DNA		CDS		Amino acids		Helix-loop-helix domain		Source
		Length (kb)	Repeat [a] (%)	Length (pb)	Identity [b] (%)	Length (aa)	Identity [b] (%)	Length (aa)	Identity [b] (%)	
Human	ENSG00000196628	413.6	23.7	2322		773	100	55		Ensembl
Chimpanzee	ENSPTRG00000010040	412.5	22.6	2259	99	752	96	56	91	Ensembl
Orangutan	ENSPPYG00000009175	361.4	20.3	2016	99	671	100	55	100	Ensembl
Rhesus	ENSMMUG00000012001	360.8	19.3	2016	98	671	99	55	100	Ensembl
Mouse	ENSMUSG00000053477	340.4	11.3	2013	92	670	98	55	100	Ensembl
Rat	ENSRNOG00000012405	340.1	11.0	1758	90	585	96	55	100	Ensembl
Dog	ENSCAFG00000000140	235.0	14.0	1857	94	618	98	55	100	Ensembl
Horse	ENSECAG00000017265	345.8	17.0	2049	96	682	98	55	100	Ensembl
Lizard	ENSACAG00000008863	268.2	3.0	1875	85	624	94	55	100	Ensembl
Platypus	ENSOANG00000003082	157.2	23.4	1647	87	548	96	55	100	Ensembl
Opossum	ENSMODG00000020455	223.0	59.7	1626	90	541	97	55	100	Ensembl
Xenopus	ENSXETG00000019086	31.4	35.4	1464	80	487	90	55	100	Ensembl
Medaka	ENSORLG00000015062	31.3	7.0	1644	71	547	72	56	98	Ensembl
Fugu	ENSTRUG00000002569	36.2	7.0	1581	71	526	75	56	94	Ensembl

Abbreviations: rhesus, rhesus macaque; xenopus, xenopus tropicalis.

[a] The proportion (%) of repetitives sequences identified by the Repeat Masker in UCSC.

[b] Identity was calculated by comparing with the human TCF4 sequence.

Supplementary Figure 1 (Appendix 1) shows an alignment of TCF4 protein sequences of 14 vertebrate species. The bHLH domain is conserved among these species as well as an element reported to play an important role the activity of the protein (Herbst and Kolligs, 2008). Only the Ensembl database annotates a longer N-terminal region that includes a signal peptide region in a human (ENSP00000381382) and chimpanzee (ENSPTRP00000040814) isoforms (amino acids position 1–31). Bioinformatic analyses were with a suite of tools, as described by Emanuelsson et al. (Emanuelsson et al., 2007) to predict the function of this N-terminal region and the subcellular localization of TCF4 in human and chimpanzee. I confirmed that both the human (amino acids position 1–31) and chimpanzee (amino acids position 1–27) sequences have predicted signal peptides (Supplementary Table I; Supplementary Figures 2–4; see Appendix 1). For the human protein, a leucine-rich nuclear export signal (amino acid position 14) was also predicted by the software NetNES (la Cour et al., 2004). Proteins having this signal can shuttle back and forth between the nuclear and cytoplasmic compartments many times during their life cycle (Pemberton and Paschal, 2005). In a recent study by Sepp et al. (Sepp et al., 2011) experimental evidence demonstrated TFC4 in fact does shuttle from the nucleus to the cytoplasm. They showed that full-length TCF4 isoforms that have longer N-termini were restricted to cell nucleus whereas other isoforms localized to both the nucleus and the cytoplasm, and found a nuclear localization signal in the protein in exons 8–9.

In addition to protein motifs, non-coding sequence nearby genes can have important regulatory functions, such as driving expression or modulating splicing, and represent an important source of information to understand gene function. These regulatory sequences can be identified by cross species DNA alignments. I used the VISTA genome browser (Frazer et al., 2004) and identified 365 ECRs (evolutionarily conserved sequences) between human and mouse with an average identity of approximately 75% (Supplementary Table II, Appendix 1). When more genetically distant species (human, dog, mouse, and opossum) were employed in a cross-species sequence comparison, all exons appear highly conserved but only a few non-exonic ECRs are present

(Supplementary Figure 5, Appendix 1). Four ECRs conserved between the human and mouse genome sequences have been tested as part of the vista enhancer project (Visel et al., 2007) (Supplementary Table III, Appendix 1). The human element hs376 (chr18:51,240,623–51,242,077 on human genome assembly hg19) is located in intron 5 of *TCF4* and was shown to drive expression of a reporter gene in the nervous system during mouse development, including the hindbrain (rhombencephalon), mid-brain (mesencephalon), and neural tube (Supplementary Figure 6, Appendix 1). Interestingly, I found reported genetic variation within this regulatory element (Supplementary Table IV, Appendix 1) and several predicted transcription factor binding sites (Supplementary Figure 7, Appendix 1) for CREB1, ZEB1, and CUX1 among others. Interestingly, CUX1 binding to the TCF4 gene promoter has been proved experimentally by Harada et al. (Harada et al., 2008) where chromatin affinity experiments were performed in HEK293 cell line. I also found four experimentally identified transcription factors binding sites from the ENCODE Chip-seq data: REST (NRSF), NFKB1, SPI1 (PU.1), and POU2F2 (Euskirchen et al., 2007; Robertson et al., 2007; Rozowsky et al., 2009) (Supplementary Figure 7, Appendix 1). Some of these transcription factors are known to be important in neurogenesis and neuronal plasticity, like REST and CREB1 (Dworkin and Mantamadiotis, 2010; Gao et al., 2011; Merz et al., 2011; Rossbach, 2011). Further study of these ECRs and the potential regulatory links revealed by growing functional genomics information might shed light on the regulation of TCF4 expression and its function in the central nervous system (CNS).

2.4 The TCF4 protein

E-proteins comprise a family of class I bHLH transcription factors homologous to the drosophila protein ‘‘daughterless’’ (Lazorchak et al., 2005; Murre, 2005) and include E12 and E47 (encoded by a single gene, E2a), HEB and TCF4 (E2-2) which form homodimers or heterodimers with other proteins in order to enable DNA binding. They have no known DNA binding activity as monomers. Class I bHLH transcription factors are expressed in a ubiquitous pattern and are able to dimerize efficiently with tissue-restricted class II factors to activate gene expression (Murre et al., 1989). They recognize a

palindromic binding site (‘‘CANNTG’’), first identified in immunoglobulin enhancers, and have a preference for C or G in the central bases. Because each partner contributes a specific DNA recognition half-site, different class I and class II heterodimers can provide distinct E box binding specificities.

Multiple alternatively spliced transcript variants that encode different TCF4 proteins have been described, with two main isoforms, E2-2A and E2-2B. E2-2B is a transcriptional repressor and a downstream target of the TCF/β-catenin complex (Kolligs et al., 2002; Petropoulos and Skerjanc, 2000) whereas E2-2A is a transcriptional regulator with diminished or absent repressor capacity (Furumura et al., 2001; Skerjanc et al., 1996). For example, the E2-2B isoform can repress the brain-specific FGF1 (fibroblast growth factor 1) by binding to an imperfect E-box in the promoter (Liu et al., 1998) while the E2-2A isoform has no activity. Likewise, interaction of E2-2B protein with MyoD can inhibit its activity by forming an inactive heterodimer (Skerjanc et al., 1996).

2.5 Biological functions and interactions of TCF4

TCF4 is widely expressed in human tissues and appears to be involved in multiple biological processes, including plasmacytoid dendritic cell development (Cisse et al., 2008; Ghosh et al., 2010; Watowich and Liu, 2010), sertoli cell development (Muir et al., 2006), B and T lymphocyte development (Bergqvist et al., 2000; de Pooter and Kee, 2010; Engel and Murre, 2001; Quong et al., 2002), thymocyte development (Wikstrom et al., 2008), the epithelial-mesenchymal transition (EMT) (Sobrado et al., 2009), the maintenance of the properties of corneal epithelial stem cells during corneal epithelial development (Bian et al., 2010; Lu et al., 2012), and brain development (see below).

In EMT, a process that allows cell migration during embryonic development and tumor invasion, TCF4 can act as an E-cadherin repressor. Functional loss of the cell–cell adhesion molecule E-cadherin is an essential event in EMT and transcriptional repression has emerged as the main regulatory mechanism. E-cadherin repression mediated by

TCF4 is indirect and independent of proximal E-boxes of the promoter (Sobrado et al., 2009). Because of the potential role of TCF4 in tumor invasion, TCF4 has been examined in cancer, including estrogen receptor negative breast cancer (Appaiah et al., 2010; Brody and Witkiewicz, 2010) and is activated in human cancers with beta-catenin deficits (Kolligs et al., 2002).

TCF4 (E2-2.1) is also an essential molecular regulator of the plasmacytoid dendritic cell (PDC) lineage (Cisse et al., 2008; Esashi and Watowich, 2009; Nagasawa et al., 2008), cells which play a central role in the defense against viruses through rapidly producing interferon I (Cervantes-Barragan et al., 2012). TCF4 shows high expression in PDCs relative to other immune cells and regulates terminal development and interferon production by controlling PDC-related genes. Impairment of TCF4 (E2-1.1) leads to a reduced number and abnormal cell surface of PDCs, which is thought to be the basis for persistent respiratory infection in patients with PHS.

TCF4 may also be involved in regulation of calcium signaling in cells through its interaction with calmodulin. TCF4 along with the E2A proteins E12 and E47 bind calmodulin in vitro at physiologically relevant calmodulin concentrations (Corneliussen et al., 1994) in an interaction modulated by sequences directly N-terminal to the basic domain. This results in inhibition of DNA binding in vitro. Increased intracellular Ca^{2+} concentration potently inhibits the transcriptional activity of TCF4 on an E-box containing reporter plasmid, as does calmodulin overexpression. Thus Ca^{2+} signaling may inhibit the transcriptional activity of E-proteins, including TCF4, by establishing an E-protein-calmodulin complex that prevents the E-protein from interaction with its target DNA (Saarikettu et al., 2004).

TCF4 expression in the brain is likely to play a major role in neurodevelopment, especially as haploinsufficiency has been associated with PHS and common variants in or near the gene confer risk of schizophrenia. However, there is a paucity of information on the role of TCF4 in the CNS. TCF4 is known to be important in the differentiation of

glial cells, especially the maturation of oligodendrocyte progenitors. In the postnatal spinal cord TCF4 expression is exclusive to oligodendrocyte lineage cells and decreases as myelination proceeds, whereas in the brain, TCF4 is also expressed in neurons, especially in the thalamus (Fu et al., 2009).

TCF4 is known to interact with several class II bHLH transcription factor genes involved in neurodevelopment: Math1, a proneural protein expressed in the differentiating neuroepithelium (Flora et al., 2007; Gohlke et al., 2008; Ross et al., 2003); HASH1, a protein necessary for the formation of distinct neuronal circuits within the CNS, especially the telencephalon (Persson et al., 2000); neuroD2, which plays important roles in neuronal differentiation and survival (Ravanpay and Olson, 2008); and Id1, a homologue of the drosophila daughterless and emc loci, which are required for correct patterning in neurogenesis (Einarson and Chao, 1995). TCF4 is also the only transcription factor showing complete co-expression with Olig2 (Fu et al., 2009), an essential regulator of ventral neuroectodermal progenitor cell fate and a requirement for oligodendrocyte and motor neuron development (Othman et al., 2011; Panman et al., 2011). Taken together, these data suggest that TCF4 is involved in an intricate combinatorial regulatory circuit during CNS development. Ravasi et al., (Ravasi et al., 2010) support a model in which the combinatorial nature of transcription factors rather than their expression in isolation is the main determinant of tissue specificity and the functionality of transcription factors complexes is associated with their conservation across species. Ravasi et al., (Ravasi et al., 2010) show that TCF4 is part of a conserved transcription factor complex of proteins co-expressed in cerebellum in human and mouse.

Math1 is highly expressed in proliferating progenitors of the rhombic lip, a specialized neuroepithelium in the developing dorsal hindbrain (Akazawa et al., 1995; Ben-Arie et al., 1996), and is also required for the differentiation of most cerebellar and pre-cerebellar structures (Ben-Arie et al., 1997; Wang et al., 2005)—for example, Math1 knockout mice lack all cerebellar granule neurons. It is important to note that cerebellar expression of TCF4 is very high (Sepp et al., 2011) and could relate to Math1. TCF4 also forms a

functional complex with HASH-1, and thus appears to be involved in the development of specific parts of the central and peripheral nervous systems (Persson et al., 2000). The HASH-1/TCF4 complex binds an E-box (CACCTG) *in vitro* and is one of the major HASH-1 interacting proteins in extracts from neuroblastoma cells. HASH-1 (MASH-1 in mouse) is a transcription factor homologous to the drosophila ‘‘achaete-scute homolog-1’’ and is essential for correct development of olfactory and most peripheral autonomic neurons, and for the formation of distinct neuronal circuits within the CNS (Persson et al., 2000).

The E proteins E12, HEB, and TCF4 all interact with neuroD2, and heterodimers bind a neuroD2 preferred E-box. These complexes have a comparable ability to induce neurogenesis, indicating a similar baseline biological activity (Ravanpay and Olson, 2008). NeuroD2 has been reported to play a critical role in the induction of neuronal differentiation, the promotion of neuronal survival, and development of thalamocortical communication (Farah et al., 2000; Ince-Dunn et al., 2006; Lin et al., 2006; Noda et al., 2006; Olson et al., 2001).

Brockschmidt et al., (Brockschmidt et al., 2011) performed knockdown of TCF4 in zebrafish embryos, which resulted in developmental delay or defects in terminal differentiation of brain and eyes. The authors conclude that TCF4-knockdown in zebrafish embryos did not seem to affect early neural patterning and regionalization of the forebrain, but may be involved in later aspects of neurogenesis and differentiation.

2.6 Human TCF4 expression

Inspection of public databases showed that *TCF4* is expressed at detectable levels across many tissues (Supplementary Figures 8 and 9, Appendix 1). However, its expression is high in prefrontal cortex, foetal brain, and BDCA4+ dendritic cells. I explored the Human Brain Transcriptome dataset, which has expression data of *TCF4* across several brain regions and developmental stages (Kang et al., 2011), and also examined the Sestan Lab Human Brain Atlas Microarrays dataset (Johnson et al., 2009) to explore the expression pattern of TCF4 in human brain. I found that *TCF4* has higher

expression in neocortex and hippocampus compared with striatum, thalamus, and cerebellum and that its expression seems to be higher at early stages of development (Figure 2.2; Supplementary Table V, Appendix 1). These brain areas have been implicated with schizophrenia by neuropsychological neuroimaging and post mortem histology and neurochemical studies (Boyer et al., 2007; Casanova et al., 2008; Harrison, 2004; Heckers, 1997; Lewis and Gonzalez-Burgos, 2008; Zierhut et al., 2010).

de Pontual et al. (de Pontual et al., 2009) found that *TCF4* is highly expressed throughout human CNS development and the sclerotomal component of the somites from Carnegie stage C13 28–32 days postfertilization (dpf). *TCF4* expression continues into adult life, expressed in lymphocytes, fibroblasts, gut, muscle, but not the heart. Sepp et al. (Sepp et al., 2011) used RT-PCR to detect *TCF4* transcripts and also found widespread expression, including testis, placenta, prostate, and in postmortem adult human brain, where they found particularly high expression, especially in the cerebellum.

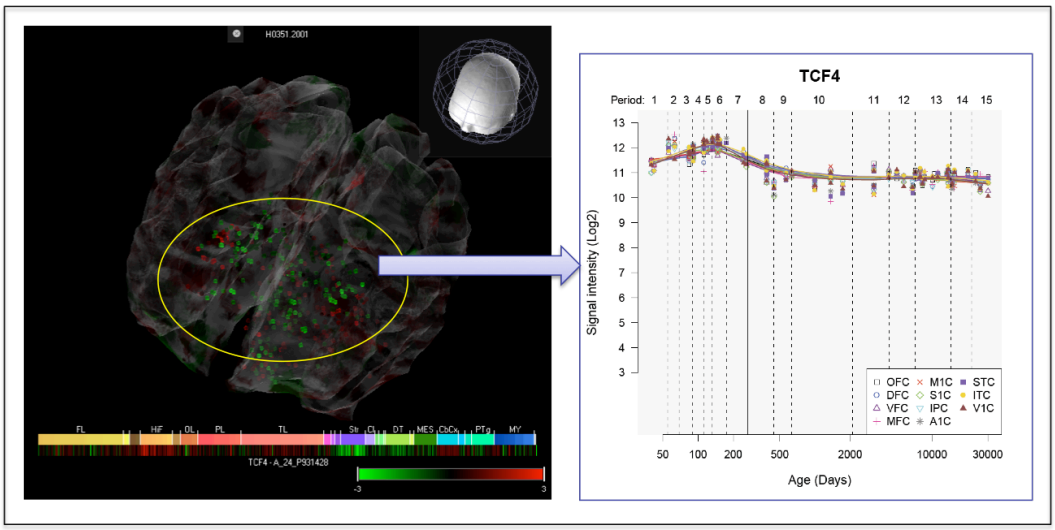


Figure 2.2 Expression of TCF4 across several brain regions and developmental stages.

Data obtained from the Human Brain Transcriptome dataset and the Sestan Lab Human Brain Atlas Microarray dataset. TCF4 has higher expression in neocortex and hippocampus compared with striatum, thalamus, and cerebellum and it expression seems to be higher at early stages of development.

2.6.1 *TCF4* RNA Transcripts

The *TCF4* gene has numerous exons and produces several mRNA transcripts. Currently two human transcripts (isoforms *TCF4A* [*ITF2A; SEF2-1A*] and the canonical sequence *TCF4B* [*ITF2B; SEF2-1B*]; accession numbers NM_001083962.1 and NM_003199.2) are commonly referred to in the literature, along with a third known isoform *SEF2-1D* (*TCF4D*). Likewise, two transcripts of human *TCF4* are annotated in the NCBI Entrez database, but four in Ensembl, 10 in UCSC and 38 in NCBI AceView database (version April 2007) (Thierry-Mieg and Thierry-Mieg, 2006). More than 459 mRNA or EST sequences are related to human *TCF4*. Forty one splice variants are recorded at Ensembl's latest version (version 66), many of which are derived from RNA-seq (*de novo* sequencing of transcripts (Ozsolak and Milos, 2011) experiments from the Illumina Body Map and ENCODE projects and therefore the splice junctions are likely to be real. Increasing availability of RNA-seq data will help to elucidate the full coding potential of the gene as well as possible transcriptional biases for different isoforms on specific tissues.

TCF4 mRNA transcripts have been experimentally verified in mice (Brzózka et al., 2010), rat (Yoon and Chikaraishi, 1994), and humans (Pscherer et al., 1996). Sepp et al. (Sepp et al., 2011) used RT-PCR in a detailed bioinformatic and experimental study of TCF4. They demonstrated the existence of several 5' exons that potentially yield 18 different N-termini in TCF4 proteins, and alternative splicing of several internal exons that may give rise to further isoforms. All the isoforms contain transactivation domain AD2 and the bHLH domain; however, only the isoforms with the longest N-terminal have the amino-terminal transactivation domain AD1. AD1 and AD2 are able to activate transcription individually and together they act synergistically (Sepp et al., 2011).

Some of these isoforms contain a nuclear localization signal (NLS) and are only found in the nucleus, whereas others are also found in the cytoplasm. Isoforms without an NLS may rely on binding to bHLH partner proteins, such as NeuroD which contains an NLS, in order to enter the nucleus. Sepp et al. (Sepp et al., 2011) also conclude that isoforms

lacking NLS can be exported from the nucleus by heterodimerization with nuclear export signal (NES) containing partner proteins. In the present study, I also detected a leucine-rich NES (amino acid position 14), which could also explain its export from the nucleus. Supplementary Table VI (Appendix 1) provides identifier mapping for isoforms across different databases that can be useful for researchers aiming to use information across these databases.

2.6.2 Potential miRNA Regulation

TCF4 is known to be regulated by at least two differentially expressed microRNAs, miR-221 and miR-222, which regulate plasmacytoid versus conventional dendritic cell development (Kuipers et al., 2010). MicroRNAs are likely to confer tissue-specific fine-tuning of *TCF4* expression, for example in neurons. In order to explore this, I used the TargetSCAN (Lewis et al., 2005) database to identify predicted miRNA binding sites in the *TCF4* UTR conserved regions. There were 45 miRNA predicted binding sites conserved across vertebrates. Of them, 37 have been associated with human diseases and 30 specifically with CNS disorders (Table 2.2). Interestingly, hsa-mir-106a/b, hsa-mir-7-1/2/3, hsa-mir-92a-1/2, hsa-mir-92b, and hsa-mir-20a/b have been associated with schizophrenia and autism (Abu-Elneel et al., 2008; Perkins et al., 2007; Talebizadeh et al., 2008). Furthermore, a genetic variant near hsa- mir-137 has been associated with schizophrenia at genome-wide significant level (Ripke et al., 2011). The targeting of hsa-mir-137 to *TCF4* has been confirmed by Kwon et al. (Kwon et al., 2013). hsa-mir-155 has been also validated experimentally according to the database mirWalk (Dweep et al., 2011) (http://www.ma.uni-heidelberg.de/ apps/zmf/mirwalk/mirnatargetpub.html).

Table 2.2 Predicted conserved miRNA target sites in *TCF4*. miRNA potentially regulating *TCF4* and associated with CNS diseases.

miRNA	Disease	Reference
hsa-mir-106a / hsa-mir-106b	Schizophrenia, Alzheimer Disease, Autistic Disorder, Glioma	(Abu-Elneel et al., 2008; Hebert et al., 2009; Lavon et al., 2010; Perkins et al., 2007)
hsa-mir-7-1 / hsa-mir-7-2 / hsa-mir-7-3	Schizophrenia, Autistic Disorder	(Abu-Elneel et al., 2008; Perkins et al., 2007)
hsa-mir-92a-1 / hsa-mir-92a-2 / hsa-mir-92b	Schizophrenia, Autistic Disorder, Medulloblastoma, Brain Neoplasms	(Nass et al., 2009; Northcott et al., 2009; Perkins et al., 2007; Talebizadeh et al., 2008; Uziel et al., 2009)
hsa-mir-20a / hsa-mir-20b	Schizophrenia, Alzheimer Disease, Medulloblastoma, Glioma, Glioblastoma	(Ernst et al., 2010; Hebert et al., 2009; Lavon et al., 2010; Northcott et al., 2009; Perkins et al., 2007; Uziel et al., 2009)
hsa-mir-93	Autistic Disorder, Glioma	(Abu-Elneel et al., 2008; Lavon et al., 2010)
hsa-mir-129-1 / hsa-mir-129-2	Autistic Disorder	(Abu-Elneel et al., 2008)
hsa-mir-363	Autistic Disorder	(Talebizadeh et al., 2008)
hsa-mir-193b	Autistic Disorder	(Abu-Elneel et al., 2008)
hsa-mir-17	Alzheimer Disease, Medulloblastoma, Glioma, Glioblastoma	(Ernst et al., 2010; Hebert et al., 2009; Lavon et al., 2010; Northcott et al., 2009; Uziel et al., 2009; Xu et al., 2010)
hsa-mir-139	Neurodegenerative Diseases	(Saba et al., 2008)
hsa-mir-155	Down Syndrome, Cerebral Hemorrhage, Stroke, Glioblastoma	(Care et al., 2007; Liu et al., 2010; Sethupathy et al., 2007)
hsa-mir-124-1 / hsa-mir-124-2 / hsa-mir-124-3	Glioblastoma, Medulloblastoma, Stroke, Cocaine-Related Disorders	(Chandrasekar and Dreyer, 2009; Laterza et al., 2009; Pierson et al., 2008; Silber et al., 2008)
hsa-mir-128-1 / hsa-mir-128-2	Neuroblastoma	(Evangelisti et al., 2009)
hsa-mir-137	Glioblastoma	(Silber et al., 2008)
hsa-mir-141	ACTH-Secreting Pituitary Adenoma	(Amaral et al., 2009)
hsa-mir-145	ACTH-Secreting Pituitary Adenoma, Atherosclerosis	(Amaral et al., 2009; Zhang, 2009)
hsa-mir-183	Glioma	(Lavon et al., 2010)
hsa-mir-200a	Meningioma	(Poethig, 2009)
hsa-mir-221	Astrocytoma	(Conti et al., 2009)
hsa-mir-367	Glioma	(Lavon et al., 2010)

2.6.3 SNPs in *TCF4*

18,636 SNPs (db130) were found at the locus, which I defined as the gene boundaries plus approximately 2 Mb in each direction to cover all proximal haplotype blocks (chr 18: 49.019.029–52.414.148, human assembly hg18). Of them, 4,417 and 6,762 have been genotyped by the HapMap and 1,000 genomes project, respectively, and 2,706 in both. Galaxy tools (Blankenberg et al., 2010) were used to intersect these SNPs with a set of putative and validated DNA functional elements which found that 12,500 SNPs (67%) are within potentially regulatory regions, including multispecies conserved sequences and known protein binding sites, and 64 were within exons (Figure 2.3). Among the latter, 29 were validated non-synonymous SNPs, as listed in the UCSC genome browser (Supplementary Table IX; Appendix 1). Williams et al., (Williams et al., 2011b) used data from the 1,000 genomes project to identify non-synonymous coding variants in the *TCF4* gene, but did not find any evidence for association and conclude that association between schizophrenia and *TCF4* is not mediated by a relatively common non-synonymous variant. They also performed allele-specific expression analysis using human post mortem brain samples, but found no evidence for *cis*-regulated mRNA expression related to genotype at the schizophrenia- associated SNP rs9960767.

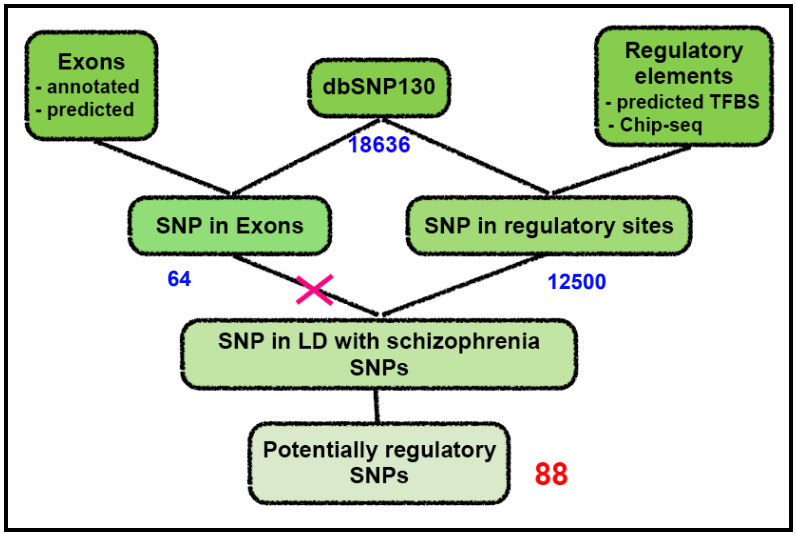


Figure 2.3 Summary of the strategy used to find the 88 potentially regulatory SNPs in LD with the SNPs associated with schizophrenia.

Eighty-eight out of the 4,417 SNPs genotyped by HapMap were both in high LD ($r^2 > 0.8$) with the SNPs associated with schizophrenia, rs4309482 and rs9960767, and within potentially regulatory sites (Table 2.3), suggesting that the disease-susceptibility alleles may capture variants influencing gene expression. To test this I looked at an eQTL dataset at http:// eqtl.uchicago.edu/Home.html (Veyrieras et al., 2008) where I found one eQTL mapping to the *TCF4* gene region, rs1261085. However, this eQTL showed low LD with the three schizophrenia-associated SNPs, for example with rs996076 it had an r^2 of 0.073, and an r^2 of 0.003 with rs4309482 and rs12966547.

Table 2.3 Putative functional SNPs in LD with schizophrenia associated SNPs

SNP 1	***SNP ID***	***Position***	***r^2***	***Functional element***
rs4309482	rs4309482	50901467	1	Nucleosome occupancy (MEC)
	rs12966547	50903015	1	phastConsElements28wayPlacMammal, phastConsElements28way, Nucleosome occupancy (MEC)
	rs4801131	50903698	1	Nucleosome occupancy (MEC)
	rs4131791	50898869	1	Nucleosome occupancy (MEC)
rs9960767	rs9960767	51306000	1	Nucleosome occupancy (MEC), Sunny RBP
	rs12327270	51319071	1	phastConsElements28wayPlacMammal, phastConsElements28way, Nucleosome occupancy (MEC), Regulatory Potential 7x, Gis Chip (c-Myc) , Sunny RBP
	rs10401120	51343496	0.867	Nucleosome occupancy (MEC), Sunny RBP

2.7 Co-expression and protein–protein interaction networks of TCF4

Network biology information was used to get an insight into the function of TCF4 by using higher-level biological information (Geschwind and Konopka, 2009). Gene co-expression networks derived from gene-expression analysis of human brains (Oldham et al., 2008) revealed 139 genes tightly co-expressed with *TCF4*. These genes were analyzed by gene-set enrichment analysis with MetaCore (GeneGO, Inc.) showing that they are characterized by specific biological processes including Alpha-2 adrenergic receptor activation of mitogen-activated protein kinase 1 and 3 (MAPK1/3) (P = 1.5 X 10^{-3}), WNT signaling pathway (P = 5.9 X 10^{-3}) and cell adhesion/synaptic contact (P = 3.0 X 10^{-3}) (Supplementary Table VII; Appendix 1). It is interesting that some of them have been previously associated with schizophrenia, such as the WNT signaling pathway (Cotter et al., 1998; Freyberg et al., 2010; Miyaoka et al., 1999) and cell adhesion/synaptic contact (Eastwood, 2004; Gross et al., 2003; Hayashi-Takagi and Sawa, 2010; O'Dushlaine et al., 2011; Stephan et al., 2009).

In addition, GeneGO Metacore (GeneGo, Inc.) manually curated information was used to identify TCF4 interacting partners, based on physical or regulatory interactions, that are also known to be associated with schizophrenia. I identified six proteins, namely TP53, NR3C1, AR, CTNNB1, ESR1, and EGFR, known to regulate or interact with TCF4 and further seven from among the co-expression partners, that is GSK3b, NRXN1, GABRA1, MILL, GRIA3, GNAL, and ADRA2A. I used the manually curated interactions and the co-expression results to build a network of potential schizophrenia susceptibility genes around TCF4 (Figure 2.4). This sheds light on its relationship with other potential schizophrenia susceptibility genes.

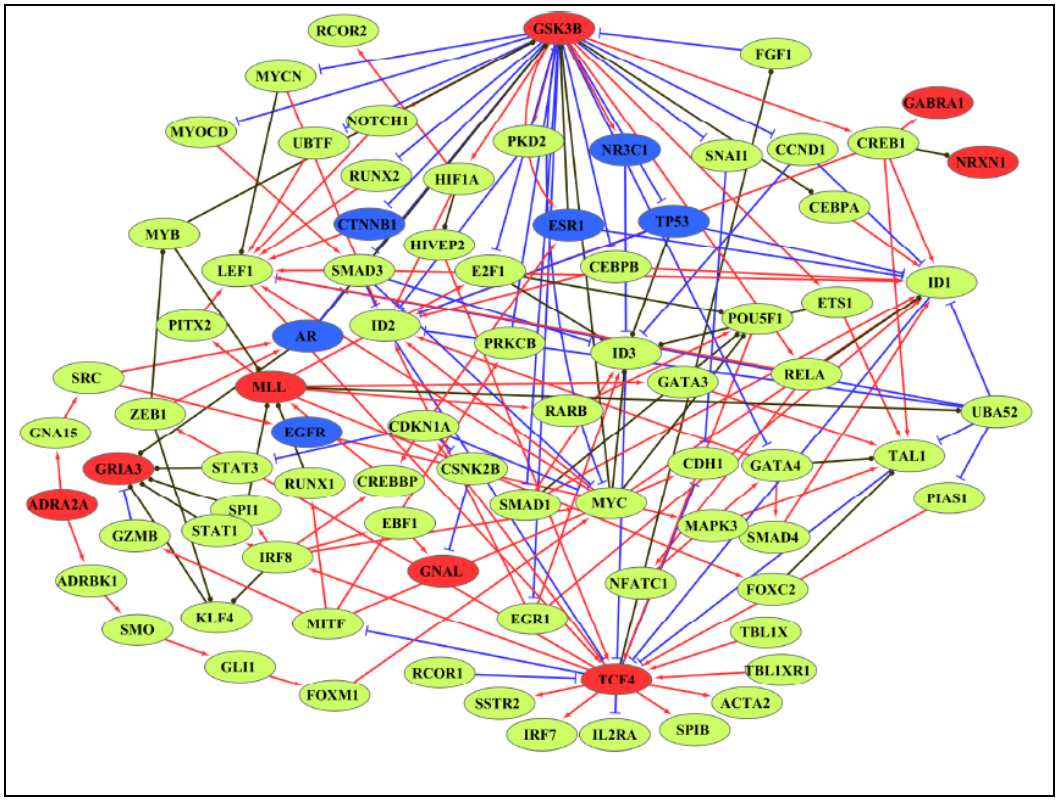


Figure 2.4 Network relationships between TCF4, genes that are coexpressed with TCF4 in human cortex and genes that are associated with schizophrenia.

Red is the genes coexpressed with TCF4 and associated with schizophrenia and blue is the genes associated with schizophrenia. The arrows show the direction of the interaction and the colour if it is activation (red), inhibition (blue), and unspecified (black).

2.8 Summary and future directions

TCF4 is still a poorly characterized gene in terms of function and biology. However, the compelling evidence for genetic association of *TCF4* with schizophrenia calls for additional detailed studies. For example, the current literature only refers to three protein isoforms at the most, but there is experimental evidence for many more (Sepp et al., 2011). Future studies will benefit from a careful dissection of the functional differences between *TCF4* splice variants, especially since they differ in functional motifs. This potential array of different protein products from the *TCF4* gene may account for its pleiotropic associations with disease.

Although the role of TCF4 during development has not been elucidated, it seems plausible, based on its known interactions with Math1, HASH1, and neuroD2, to suggest that TCF4 plays an important role during CNS development. Because of its interaction with calmodulin and involvement in calcium signaling, TCF4 may also be involved in acute functional effects in neurons. The advent of second generation sequencing technologies will allow a more thorough study of transcription factor binding site targets of *TCF4*, using methods such as ChIP-seq, as well as isoform abundance at different points during development of CNS tissues using RNA-Seq. These analyses will reveal further information on the TCF4 biological regulatory network, the role of the protein in brain development and function and, in turn, TCF4's involvement in psychiatric and neurodevelopmental disorders.

The role of TCF4 in brain neural function, especially in relation to schizophrenia, also needs to be elucidated. For example, preliminary studies are examining the role of *TCF4* polymorphisms on putative endophenotypes of schizophrenia have shown an effect on sensory gating, as measured by the P50 suppression of the auditory evoked potential (Quednow et al., 2012), and on prepulse inhibition of the acoustic startle response, a different measurement of sensorimotor gating (Quednow et al., 2011). These studies may provide important information on the relationship between *TCF4* risk polymorphisms and neuropsychiatric dysfunctions in schizophrenia.

Examination for putative functional genetic variants revealed that several common SNPs are both within likely functional DNA elements and in LD with SNPs associated to schizophrenia. In line with these findings, one of these elements is known to drive gene expression in mouse CNS. Additional studies exploring the function of these DNA sequences can provide clues as to which transcription factors are regulating TCF4 and in turn how TCF4 is being integrated into wider regulatory processes. In addition, *TCF4* transcripts have several potential binding sites for miRNAs whose expression has been associated with mental illnesses. Additional validation of these regulatory links is warranted and could provide important clues between *TCF4* and other genes underlying risk of mental disorders. Thus, the information compiled here suggests it may be possible to use additional targeted functional studies to start drawing a functional network relating TCF4 to other key molecular players in brain development and mental illness. Furthermore, I showed that by compiling gene-expression data and protein–protein interaction information TCF4 is readily linked to other genes associated with schizophrenia. A major challenge will be to understand the mechanisms relating these genes and in which context, for example cell type or developmental stage, their interactions are important as mediators of risk for schizophrenia. This effort will need to be undertaken by a wide community applying a plethora of experimental approaches. Nonetheless, this represents the start of a path lead to biological understanding and possibly new treatments for schizophrenia and other psychiatric disorders.

2.9 Methodology.

2.9.1 Sequence analyses

The annotated *TCF4* gene was searched in Ensembl data base version 56 (http://www.ensembl.org) and UCSC genome browser (http://genome.ucsc.edu/). Ensembl annotates 47 orthologues for human *TCF4* (ENSG00000196628) gene, but the analysis was restricted to only 14 representative species. The species used were: Human (specie *Homo sapiens*, assembly GRCh37), Chimpanzee (*Pan troglodytes*, CHIMP2.1), Orangutan (*Pongo pygmaeus*, PPYG2), Macaque (*Macaca mulatta*, MMUL 1.0), Mouse (*Mus musculus*, NCBI m37), Rat (*Rattus norvegicus*, RGSC 3.4), Dog (*Canis familiaris*, CanFam 2.0), Horse (*Equus caballus*, Equ Cab 2), Anole Lizard (*Anolis carolinensis*, AnoCar1.0), Platypus (*Ornithorhynchus anatinus*, Ornithorhynchus_anatinus-5.0), Opossum (*Monodelphis domestica*, monDom5), Xenopus (*Xenopus tropicalis*, JGI 4.1), Medaka (*Oryzias latipes*, HdrR) and Fugu (*Takifugu rubripes*, FUGU 4.0).

All the analyses were performed with the canonical protein, which Ensembl defines as the longest deduced transcript for each species. Manual inspection of the predicted orthologous revealed a possible error. The predicted orthologous in Dog is ENSCAFT00000000229 but this gene only has 3 exons, compared with 21 in human. A predicted paralogous of this gene (ENSCAFT00000000232) has 17 exons. Because an intron loss event is needed to explain this and the more parsimonious answer is that orthologous of the human TCF4 is ENSCAFT00000000232 and that ENSCAFT00000000229 may have been generated by a retrotransposition event, which could explain the fewer introns of this gene. ENSCAFT00000000232 was retained as the orthologous for further analyses.

Sequence identity and similarity was calculated with BLAST (Altschul et al., 1990). DNA sequence repeats identified by Repeat Masker were extracted from the UCSC tables using the software Galaxy (Blankenberg et al., 2010). Amino-acid alignments were performed with MUSCLE (Edgar, 2004) and the visualization was generated with Clustal

X (2.0.12) (Larkin et al., 2007). The secondary structure analysis was performed using the program ali2d (Biegert et al., 2006) (http://toolkit.tuebingen.mpg.de/ali2d). In order to characterize the N- terminal region found in chimpanzee and human I followed the protocol described by Emanuelsson et al., (Emanuelsson et al., 2007), which involves the use of a suite of bioinformatic software publicly available to predict subcellular location of eukaryotic proteins.

2.9.2 TCF4 transcripts

The alternative transcripts of the human TCF4 gene were obtained from NCBI AceView (version April 2007) (Thierry-Mieg and Thierry-Mieg, 2006) (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html), Ensembl (release GRCh37) and UCSC (Mar. 2006 (NCBI36/hg18) assembly). Next TCF4 expression was examined in different tissues using the microarray data integrated in Geneinvestigator (V3, http://www.geninvestigator.ethz.ch/) (Hruz et al., 2008), BioGPS (http://biogps.gnf.org/#goto=welcome) (Wu et al., 2009) and the Sestan Lab Human Brain Atlas Microarrays data set, which is available at the UCSC Browser.

2.9.3 Analysis of SNPs in the TCF4 gene

Galaxy (Blankenberg et al., 2010) was used to extract all SNPs annotated in db130 within the *TCF4* locus (chr. 18: 49019029 – 52414148, assembly hg18) and a bioinformatics protocol was applied to evaluate if they were within possible functional sites. SNPs were also extracted from HAPMAP (Release #28) and 1000 genomes pilot data 1 (Durbin et al., 2010). Functional regions were defined with several UCSC Tables (Supplementary Table 8). The software SNAP (Johnson et al., 2008) was used to evaluate the LD between SNPs.

2.9.4 Identification of possible regulatory elements

Multispecies Conserved Sequences (MCS) within the *TCF4* locus identified with VISTA genome browser (Frazer et al., 2004) and UCSC tracks were examined to investigate the possible function of these regions based on the overlap with functional genomic

experiments results deposited in the browser. I searched for conserved miRNA predicted binding sites on *TCF4*'s 3' UTR using TargetScan DB (Lewis et al., 2005). Predicted miRNA were examined to determine their association with human diseases using the miRNA and Disease database (Lu et al., 2008).

2.9.5 Co-expression and protein-protein interactions networks of TCF4

Gene co-expression data was taken from a published work for human brain (Oldham et al., 2008) to identify TCF4 co-expression partners. These were analysed using gene-set enrichment analysis with MetaCore (GeneGO, Inc) and Cytoscape (Cline et al., 2007).

Chapter 3 Effect of *TCF4* knockdown on neuronal progenitor cells proliferation

3.1 Introduction

In addition to the literature review presented on Chapter 2, here I provide additional information regarding the molecular and cell biology of TCF4 gathered post-publication of our literature review (Navarrete et al., 2013). New studies have provided additional support to the statistical association between genetic variation at the *TCF4* locus and schizophrenia (Aberg et al., 2013; Albanna et al., 2014; Irish Schizophrenia Genomics and the Wellcome Trust Case Control, 2012; Kwon et al., 2013; Ripke et al., 2013), other studies have examined the association between *TCF4*'s expression and schizophrenia (Umeda-Yano et al., 2014; Wirgenes et al., 2012), and additional studies have investigated the relationship between risk variant at *TCF4* locus and cognition (Albanna et al., 2014; Zhu et al., 2013). However, there has been very little progress in understanding the molecular and cellular mechanisms that might mediate the association between *TCF4* variation and risk for schizophrenia. To date, only one study has used manipulation of *TCF4* expression to investigate its effect in human cells of neuronal origin (Forrest et al., 2013). Knockdown of *TCF4* in human neuroblastoma-derived cells (SH-SY5Y) induced changes in the expression of genes involved in TGF-b signalling, epithelial to mesenchymal transition, neuronal differentiation and apoptosis. However, the use of the SH-SY5Y cell line as neuronal cell model has been criticised, among other things due to its chromosomal abnormalities which are expected of cancer cell lines (Bray et al., 2012).

In an unpublished study conducted by colleagues in Dr Bray's group the expression of *TCF4* was manipulated in two karyotypically-normal human neuronal progenitor cell lines (CTX0E16 and CTX0E03) using RNAi and the effects on the cellular transcriptome was assessed using expression micro-arrays. The gene-expression data for the CTX0E03 cell line showed nominally significant ($P < 0.05$) gene expression differences shared by the

two *TCF4* siRNA conditions in the same direction relative to control at 628 probes while for CTX0E16 cell line the same was observed for 197 probes. Interestingly, these differentially expressed probes were enriched for genes involved in cell cycling in both cell lines. However, weighted gene co-expression network analyses (WGCNA, (Langfelder and Horvath, 2008)) showed that the module enriched for cell cycling was positively correlated with TCF4 expression in one cell line and negatively correlated in the other. This suggests that TCF4 knockdown might have cell-specific effects on cell cycling. The implication that knockdown of TCF4 can affect cell proliferation in neuronal progenitors prompted the study presented in this Chapter.

TCF4 is a member of the class I basic helix-loop-helix (bHLH) family of proteins and previous studies have shown that TCF4 is highly expressed in the brain (de Pontual et al., 2009; Sepp et al., 2011). The bHLH family of transcription factors have been shown to be essential in determining the cell fate of neuronal progenitor populations and animal studies have indicated that it participates in the regulation of proliferation, differentiation and migration of distinct neuronal progenitors during neurogenesis (Ben-Arie et al., 1997; Bhattacharya and Baker, 2011; Chiaramello et al., 1995; Gohlke et al., 2008; Olson et al., 2001; Pagliuca et al., 2000; Ross et al., 2003). Outside the CNS, TCF4 appears to be involved in multiple biological functions including in the immune system where it plays an important role in lymphocyte development (Bergqvist et al., 2000; de Pooter and Kee, 2010; Engel and Murre, 2001; Murre, 2005; Quong et al., 2002; Zhuang et al., 1996) and regulates plasmacytoid dendritic cells (Cisse et al., 2008; Ghosh et al., 2010; Watowich and Liu, 2010).

TCF4 expression in the brain is likely to play a key mechanistic role in neurodevelopment by interacting with several transcription factors involved in defining neuronal functions. These factors include those from the class II bHLH family of transcription regulators like Math1 (Flora et al., 2007), HASH1 (Persson et al., 2000), neuroD1 and neuroD2 (Brzózka et al., 2010; Ravanpay and Olson, 2008), and the inhibitor factors Id1 and Id2 (Einarson and Chao, 1995; Langlands et al., 1997; Parrinello et al., 2001), which are essential to

determine the fate of cells in the mammalian nervous system and are considered essential to specify all the neurons of the mammalians nervous system (Bhattacharya and Baker, 2011).

The role of TCF4 during development is supported by findings from human and animal studies. In humans, *TCF4* deletions leading to haploinsufficiency are a confirmed cause of Pitt-Hopkins syndrome, a severe neurodevelopmental disorder (Amiel et al., 2007; Whalen et al., 2012). Findings in animal models have confirmed that TCF4 is critical for normal brain development. Homozygous *Tcf4* knockout mice die soon after birth and present with disrupted pontine nucleus development (Flora et al., 2007), while TCF4 knockdown in zebrafish results in developmental delay or defects in terminal differentiation of brain and eyes (Brockschmidt et al., 2011). Transgenic mice overexpressing *Tcf4* post-natally present with cognitive and sensorimotor impairments (Brzózka et al., 2010).

TCF4 is a downstream target of the WNT/TCF pathway where its expression is regulated by β-catenin via its TCF binding site in cancer cell lines (Kolligs et al., 2002) (Kollings et al 2002). The WNT signalling pathway is important in neuronal development (Abu-Khalil et al., 2004) and plays an essential role in several cellular processes including regulating cell proliferation, differentiation and neurogenesis (Chevallier et al., 2005; Malaterre et al., 2007; Zechner et al., 2003). TCF4 has been identified as a direct target of the imprinted gene *Zac1/Plagl1* during early neurogenesis in mouse neural stem cells (Schmidt-Edelkraut et al., 2014). Zac1 is a zinc finger protein expressed in the brain described as a regulator of proliferation, cell cycle arrest and apoptosis (Liu et al., 2011; Pagotto et al., 1999; Spengler et al., 1997; Valente et al., 2005). Schmidt-Edelkraut et al, (Schmidt-Edelkraut et al., 2014) also showed that TCF4 regulates the expression of the cycling dependent kinase inhibitor gene $p57^{kip2}$, increasing the number of cells in G1 phase. They conclude that cycle arrest in mouse neuronal progenitors controlled by Zac1 occurs through the induction of $p57^{kip2}$ via TCF4. Additionally, TCF4 participates in the induction of cell cycle arrest by regulation of $p21^{cip1}$ gene expression in colorectal cancer cells lines

(Herbst et al., 2009). $p21^{cip1}$ is a negative regulator of cell proliferation and a direct target of Zac1 (Liu et al., 2011; Spengler et al., 1997).

Overall, the gene expression changes observed by Dr Bray's group following *TCF4* knockdown in human neural progenitor cells together with previous evidence from human and animal model cell biology and biochemical studies lend support to the idea that TCF4 has a role in cell proliferation. Therefore, I tested whether *TCF4* knockdown would lead to changes in the proliferation of the CTX0E03 and CTX0E16 neural progenitor cell lines and, as predicted by the micro-array results, this effect would differ between them.

3.1.1 The aim of the present study

The aim of the experiments presented in this chapter was to test the hypothesis that *TCF4* knockdown leads to changes in the proliferation of the CTX0E16 and CTX0E03 neural progenitor cell lines. In addition, in accordance with the WGCNA data, I tested the hypothesis that this effect differs in direction between the cell lines. To this end, I used RNAi to knockdown endogenous *TCF4* in the two neural progenitor cell lines followed by a cell proliferation assay.

3.2 Methodology

3.2.1 Human neural progenitor cell line

Experiments were carried out in two clones (CTX0E03 and CTX0E16) of a conditionally immortalised neural progenitor cell line. This cell line was obtained from ReNeuron Ltd (www.reneuron.com) under a Material Transfer Agreement. It was derived from first trimester human foetal cortical neuroepithelium (Pollock et al., 2006). The cells have been conditionally immortalised by genomic incorporation of the *c-myc*ER^{TAM} transgene, which can be conditionally activated by 4-hydroxy-tamoxifen (4-OHT) to stimulate proliferation (Littlewood et al., 1995). A detailed description of the CTX0E03 cell line can be found in Pollock et al. (Pollock et al., 2006).

Cells were grown in laminin-coated Nunclon^{TM} Surface T75 culture flasks (Thermo Scientific, MA, USA) in Reduced Modified Media (RMM) at 37 ºC and 5% CO_2 in a humidified incubator. Nunclon^{TM} Surface T75 culture flasks (Thermo Scientific, MA, USA) were coated with Engelbreth-Holm-Swarm murine sarcoma basement membrane laminin (Sigma Aldrich, Dorset, UK) at 1 µg/cm^2. For T75 flask, 75 µL of laminin was dissolved in 7.5 mL of Dubelcco's Phosphate Buffered Saline (DPBS) pH 7.4 (Sigma-Aldrich). Flasks were incubated for 3 hours at 37 ºC and then rinsed with warm DPBS.

Cells were expanded in 10 mL of Reduced Modified Media (RMM). RMM media was prepared by supplementing DMEM: F12 with 15 mM HEPES and sodium bicarbonate but without L-glutamine (Sigma-Aldrich, D6421) with the following components: 0.03% human serum albumin, 2 mM L-glutamine, human transferrin (100 µg/mL), putrescine dihydrochloride (16.2 µg/mL), human insulin (5 µg/mL), human progesterone (60 ng/mL), sodium selenite (40 ng/mL), 100 nM 4-OHT (all Sigma-Aldrich), and the growth factors bFGF (10 ng/mL) and hEGF (20 ng/mL) (Peprotech, Inc., NJ, USA). Table 3.1 shows in detail concentrations and volumes used for each component.

Cells were grown to 70-90% confluence and the media was changed every two days. When cells had expanded to the required confluence they were passaged using Accutase

solution (Invitrogen). 3 mL of accutase was added to the flask and incubated for 5-10 minutes at 37 ºC. Following this, the cells were resuspended in RMM to inactivate the accutase and transferred to a 15 mL Falcon tube. For cell counting, 10 µL of the cell suspension was placed on a haemocytometer and counted under a light microscope. To pellet the cell, the tube was centrifuged at 900 rpm for 5 minutes at room temperature. The pelleted cells were then resuspended at an appropriate density in RMM media.

Table 3.1 Components of Reduced Modified Media (RMM).

Material Names	**Stock concentration**	**Final concentration**	**Supplier**	**Volume for 500 mL**
DMEM:F12	-	-	Sigma: D6421	500 mL
Discard volume				10.5 mL
Human serum albumin solution	20%	0.03%	PAA: C11-096	0.75 mL
Human apotransferrin	50 mg/mL	100 µg/mL	SCIPAC: T100-5	1.0 mL
Putrescine dihydrochloride	8.1 mg/mL	16.2 µg/mL	Sigma: P5780	1.0 mL
Human recombinant insulin	10 mg/mL	5 µg/mL	Sigma: I9278	0.25 mL
Progesterone	30 µg/mL	60 ng/mL	Sigma: P6149	1.0 mL
L-glutamine	200 mM	2 mM	Sigma:G7513	5.0 mL
Sodium selenite	20 µg/mL	40 ng/mL	Sigma: S9133	1.0 mL
Human FGF-basic	10 µg/mL	10 ng/mL	PeproTech: 100-18B	0.5 mL
Human EGF	100 µg/mL	20 ng/mL	PeproTech: AF-100-15	100 µL
4-OHT	1 mM	100 nM	Sigma: H7904	50 µL

3.2.2 siRNA transfection

CTX0E16 and CTX0E03 cells used for transfection were expanded in laminin-coated T75 flask to 90-100% confluence. Twenty-four hours before siRNA transfection cells were passaged using Accutase solution (Invitrogen), $1*10^6$ cells were seeded per T75 flask that would be used for all siRNA conditions. For proliferation assays 4500 cells were cultured per well in a Nunc-Immuno™ MicroWell™ 96 well solid plate. All cells were cultured, on laminin, in 10 mL of RMM media without 4-OHT (represented: RMM ++-). 4-OHT was excluded so that proliferation was not artificially stimulated by *c-myc* overexpression. CTX0E16 and CTX0E3 cells can retain their nestin-positive, neuronal progenitor state in the presence of growth factors bFGF and EGF (Dr G. Anderson and Dr B. Williams, personal communication). Separate flasks of CTX0E16 and CTX0E03 cells were transfected with either Stealth siRNA1, Stealth siRNA2, Stealth siRNA™ negative universal control and positive control BLOCK-iT™ Alexa Fluor® Red Fluorescent oligo (all from Invitrogen).

For proliferation assays 10 wells, containing CTX0E16 or CTX0E03 cells, were transfected for each siRNA treatment. Transfection of siRNAs was performed using the N-TER™ Nanoparticle siRNA Transfection System (Sigma). RNAi was achieved using pre-designed Stealth RNAi™ (Invitrogen), which uses next-generation RNAi chemistry to prevent off-target effects. Two non-overlapping Stealth RNAi™ siRNA for *TCF4* (siRNA1: S13863 and siRNA2: S13864) were transfected alongside a Stealth RNAi™ siRNA negative universal control (Invitrogen), and a BLOCK-iT™ Alexa Fluor® Red Fluorescent Oligo to assess the transfection efficiency. The sequences of the siRNA targeting *TCF4* are shown in Table 3.2. siRNA1 targets exon 18 and siRNA2 targets exon 15 of full length *TCF4* mRNA sequence (NM001243226.1).

Table 3.2 Details of siRNAs used for *TCF4* knockdown.

siRNA ID	Oligo	Sequence	*TCF4* exon
siRNA1	S13863	GCUCUGAGAUCAAAUCCGA	18
siRNA2	S13864	GAAGGACCCUUACACUCUU	15

All siRNA and the red fluorescent oligo were combined with N-TER™ transfection reagent to a 9 nM final concentration in 12 mL RMM ++-. Details of reagents used per each transfected flask are shown in Table 3.3. In separate microcentrifuge tubes siRNAs and the N-TER peptide were mixed by vortexing and incubated at room temperature for 30 minutes, allowing the siRNA and NTER peptide to combine. For each transfection, 166 µL of the siRNA/N-TER or RedOligo/N-TER mix was added to a falcon tube containing 12 mL of RMM ++- and mixed by inversion. Media from previously prepared flasks and wells was replaced with RMM++- containing siRNA or Red Fluorescent Oligo. 10 mL of the mix was added to each flask and 50 µL to each well.

Table 3.3 Component of one siRNA transfection reaction

Reagent	Concentration	Final volume added
CTX0E16 or CTX0E03 cells:		
Per flask	$1*10^6$cells	-
Per well	4500 cells	-
2 nM - 2 µM siRNA:	20 µM	
siRNA1		6.5 µL
siRNA2		6.5 µL
negative control siRNA		6.5 µL
Red Fluorescent Oligo		6.5 µL
Dilution buffer	-	93.5 µL
N-TER	-	16 µL
H_2O	-	84 µL
Total	-	166 µL

Twenty-four hours post-transfection, its efficiency was estimated by monitoring fluorescence of cells transfected with BLOCK-iT™ Alexa Fluor® Red Fluorescent Oligo by fluorescence microscopy.

Media was replaced at 48 hours post-transfection with RMM (++-). At 90 hours post-transfection cells from each T75 were separately harvested by addition of Accutase. Resuspended cells from each treatment were divided into 2 Falcon tubes and centrifuged at 900 rpm for 5 minutes. One of the cell pellets was used for RNA extraction to which 500 µL of Tri-Reagent® Solution (Ambion, Life Technologies, Grand Island, N.Y.) was added. The second pellet was used for protein extraction. Both tubes were stored at -80 °C.

Nunc-Immuno™ MicroWell™ 96 well solid plates containing cells transfected with siRNA1, siRNA2 and Stealth siRNA™ negative universal control were washed with 100 µL per well of warm DPBS, and then fixed with 4% paraformaldehyde (PFA) for further proliferation assays. A detailed protocol for the proliferation assays is described below in this section.

An overview of the transfection time-line is represented in Table 3.4. The whole protocol for the 4 treatments (siRNA1, siRNA2, Stealth siRNA™ negative universal control and BLOCK-iT™ Alexa Fluor® Red Fluorescent oligo) was repeated five times.

Table 3.4 Time-line protocol for *TCF4* knockdown assay.

Day -1	**Cells are seeded on laminin-coated flasks with 10 mL of RMM ++-, or in laminin-coated wells of a 96 well solid plate with 50 µL RMM ++-**
Day 0	Transfect: • 1 flask with siRNA 1 • 1 flask with siRNA 2 • 1 flask with negative control siRNA • 1 flask with red oligo • 36 wells of a 96 well plate, 12 wells for each siRNA condition (siRNA1, siRNA2 and negative control siRNA)
Day 1	Check fluorescence of cells transfected with red-oligo
Day 2	Replace media to all flasks and 96 well plate with RMM ++- (10 mL per flask and 50 µL per well)
Day 3	-
Day 4	Collect cells from the different treatments for RNA extraction and protein extraction. Fix cells in proliferation plate with 4% PFA.

3.2.3 Total RNA extraction

Pelleted cells were separately homogenized with 500 µL of Tri-Reagent® Solution (Ambion, Life Technologies, Grand Island, N.Y.). After homogenization, samples were transferred to a clean microcentrifuge tube. 50 µL of 1-Bromo-3-cloropropane (Sigma-Aldrich) was added to each sample, vortexed for 2 seconds and incubated at room temperature for 10 minutes. The samples were then transferred to a Phase lock gel tubes (5Prime) and centrifuged at 12000 rpm for 15 minutes at 4 ºC. The upper aqueous phase was transferred to a clean microcentrifuge tube and 250 µL of isopropanol (Sigma-Aldrich) were added. The tubes were vortexed for 5-10 seconds, incubated at room temperature for 10 minutes and centrifuged at 12000 rpm for 15 minutes at 4 ºC. The supernatant was discarded and the pellet was washed with 500 mL of 75% ethanol (Sigma-Aldrich) and centrifuged at 7500 rpm for 5 minutes at room temperature. The

ethanol was removed and the pellet was air dried before being resuspended in 30 µL of nuclease-free water (Sigma-Aldrich) and stored at -80 ºC.

3.2.4 DNAse treatment of total RNA samples

Extracted RNA samples were treated with TURBO DNA-*free*™ (Ambion) to remove any contaminating genomic DNA. 30 µL of total RNA from each sample was treated with 3.5 µL of 10x TURBO DNAse Buffer and 2.1 µL of TURBO DNAse (2 U/µL). The samples were gently mixed and then incubated at 37 °C for 30 minutes. The samples were then incubated at room temperature with 7 µL of DNAse Inactivating Reagent (Ambion) for 2 minutes, while being mixed occasionally. Finally, the samples were centrifuged at 1000 rpm for 2 minutes and the supernatant was transferred to a clean 0.6 mL microcentrifuge tube and stored at -80 °C.

3.2.5 RNA quantification

A NanoDrop™ 1000 spectrophotometer was used to measure the concentration and purity of each RNA sample. 1 µL of each sample was used to quantify the absorbance of RNA at 260 nm. The presence of contaminating proteins or phenol was assessed with the ratio 280/260 nm, while the ratio 260/230 nm was used to determine the presence of other contaminants.

3.2.6 Reverse Transcription (RT)

RT reactions were performed using SuperScript™ III reverse transcriptase (Invitrogen) and random decamers (Ambion). SuperScript™ III is an engineered version of M-MLV reverse transcriptase that is more efficient for synthesizing cDNA. It has a reduced RNase H activity and increase thermal stability. RT reactions were performed for each sample using 1 µg total RNA concentration in a final reaction volume of 20 µL. Table 3.5 shows the volumes and concentration of reagents used. The RT was carried out in two steps. First, the RNA was incubated at 65 °C for 5 minutes with the random decamers, dNTP mix (10 mM) (Promega) and nuclease-free water (Sigma-Aldrich). Then, the samples were placed on ice for 1 minute. On the second step, buffer, DTT, SuperScript™

III RT and nuclease-free water was added. The samples were transferred to a G-storm thermal cycler (Somerston Biotechnology Centre, UK) for the following incubation steps: 25 ºC for 5 minutes, 42 ºC for 1 hour, 50 ºC for 30 minutes, 55 ºC for 30 minutes and a final temperature step at 70 ºC for 10 minutes was used to inactivate the enzyme.

Table 3.5 Details of reagents used for each RT reactions.

Reagent	Supplier	Concentration	Volume (µL)
Random decamers primers	Ambion	50 µM	2
dNTP mix (dATP, dTTP, dCTP, dGTP)	Promega	10 mM	1
Nuclease-free water	Sigma	-	9
First-Strand Buffer (250 mM Tris-HCL, 375 mM KCL, 15 mM $MgCl_2$)	Invitrogen	5x	4
DTT		0.1 M	1
SuperScript III RT		200 U/µL	1
RNA template	-	1 µg/µL	1
Total	-	-	20

The resulting cDNA samples were diluted 1/7, by adding 120 µL of nuclease-free water (Sigma-Aldrich) to each 20 µL of cDNA. This concentration is referred to in this thesis as the "standard cDNA concentration". cDNA samples were stored at -20 ºC.

3.2.7 Quantitative PCR (qPCR)

Quantitative PCR was used to assess *TCF4* knockdown. qPCRs were completed on 96-well plates using an MJ Chromo4™ Real Time PCR Detector (Bio-Rad) and MJ Opticon Monitor analytic software (Bio-Rad). Reactions were carried out in a total volume of 20 µL, containing diluted cDNA, primers at 200 nM and 1x HOT FIREPol® EvaGreen® qPCR Mix Plus (Solis BioDyne). Table 3.6 shows the volume and concentration of the reagents used. The cycling programme used for qPCR reactions can be found in Table 3.7.

Table 3.6 Details of reagents used for each qPCR reaction.

Reagent	Supplier	Concentration	Volume (µL)
HOT FIREPol ® EvaGreen ® qPCR Mix Plus	Solis BioDyne	5x	4
Forward and reverse primer mix	Sigma-Aldrich	2 µM	2
Nuclease-free water		-	6
cDNA template/nuclease-free water	-	-	8
Total	-	-	20

Table 3.7 Typical qPCR cycling conditions.

Step	Temperature (°C)	Time (minutes)	Number of Cycles
Initial denaturation	95	15	1
Denaturation	95	0.3	44
Annealing	60	0.3	
Extension	72	0.3	
For the melting profile, the temperature was raised from 60 to 95 and the fluorescence read every 1°C increase in temperature with a hold time of 10 sec.			

QPCR primers were designed using the Primer3Plus program available online (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). Sequences of the primers used for qPCR are shown in Table 3.8.

Table 3.8 Detail of primers sequences used for qPCR analysis.

Gene	Primer	Sequence (5' - 3')
TCF4	Forward	GAAAGCTGCGTGTCTGAAAA
	Reverse	CATCTGTCCCATGTGATTCG
GAPDH	Forward	CCTGACCTGCCGTCTAGAAA
	Reverse	ATCCTGGTGCTCAGTGTAGCC
RPL13A	Forward	CCTGGAGGAGAAGAGGAAAGAGA
	Reverse	TTGAGGACCTCTGTGTATTTGTC
HPRT1	Forward	TGACACTGGCAAAACAATGCA
	Reverse	GGTCCTTTTCACCAGCAAGCT

Three technical replicates for qPCR reactions were performed to measure *TCF4* and the internal control genes *RPLI3A*, *GADPH* and *HPRT1* for each cDNA sample. QPCRs for *TCF4* and the internal controls were carried out on the same PCR plate for every transfection experiment and each treatment (siRNA1, siRNA2 and Stealth siRNA™ negative universal control). The level of expression of each gene was measured against a relative standard curve created by serial dilution of pooled cDNA taken from each of the assayed samples. The following arbitrary values were given to the standard curve: "5000" (undiluted), "1000", "200", and "40", according to their relative dilutions. Consequently, a relative value was obtained for each of the three replicate reactions for each cDNA sample.

3.2.8 Statistical analysis for qPCR

Description of the data:

1. Transfection experiments were done in five biological replicates. Each biological replicate has three treatments, siRNA1, siRNA2 and Stealth siRNA™ negative universal control, leading to five biological replicates per treatment.
2. cDNA was obtained from RNA for the five biological replicates for each treatment. This gave 15 cDNA samples in total.

3. qPCR was used to measure the expression of *TCF4* and three internal control genes, *RPLI3A*, *GADPH* and *HPRT1*. This was performed in triplicate, leading to three technical replicates for each of the 15 cDNA samples. qPCR for *TCF4* and the internal controls was carried out on the same PCR plate.
4. For each technical replicate I obtained one normalized value for *TCF4* per treatment per biological replicate. Normalization was calculated using the geometric normalization method (Vandesompele et al., 2002). Thus, from each technical replicate I obtained one normalized value for *TCF4* for each treatment per biological replicate. In total I obtained three technical replicates for 5 biological replicates each including the three treatments. These normalized values were used for statistical analyses.

The objective of the statistical analyses was to identify expression changes on *TCF4* mRNA level between control siRNA and siRNA conditions. To this end I used a linear mixed model including treatments as a fixed effect and biological replicates as random effects. I used a linear mixed model because it considers the systematic variation between biological replicates. Statistical analysis was performed using a linear mixed model in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.) Differences of $P < 0.05$ were considered significant. The exact command using R was:

lmer(log(normalized expression) ~ 1 + treatment + (1|biological_replicate),data=qpcr.m).

3.2.9 Proliferation assay

Cell fixing:

Nunc-Immuno™ MicroWell™ 96 well solid plates containing cells transfected with siRNA1, siRNA2 and Stealth siRNA™ negative universal control were washed with 100 µL per well of warm DPBS, and then fixed with 4% PFA for 20 minutes at room temperature. This was followed by two washes with DPBS for 5 minutes. Plates were stored at 4 °C with 100 µL per well of DPBS.

Immunostaining:

Plates were incubated with 2N HCl, 50 µL/well for 25 minutes. Next, HCl was neutralized with 100 µL per well of 0.1 M Borate buffer for 10 minutes at room temperature. The wells were washed twice with 100 µL DPBS and then incubated with 50 µL/well of blocking solution (0.3% triton X-100, 5% normal donkey serum in DPBS) for 1 hour at room temperature. Following this, the plates were incubated overnight at with 4 °C with primary antibody for anti-Ki67 (Abcam, ab15580) diluted in blocking solution 1:500. The next day the primary antibody was removed and the wells were washed twice with DPBS. Subsequently, the plates were incubated with 50 µL/well of blocking solution for 30 minutes at room temperature and the secondary antibody, Alexa Fluor® 555 Donkey Anti-Rabbit IgG (Invitrogen, A31542), diluted 1:500 in blocking solution was added. The secondary antibody was incubated for 2 h at room temperature covered from light. Next, the cells were washed two times with DPBS and then DAPI diluted 1:1000 in DPBS was added for 5 min at room temperature. The cells were then washed with DPBS and the plates scanned using the CellInsight™ Personal Cell Imager Cytometer (Thermo Scientific).

Analysis of cell proliferation using Ki67 in CellInsight

The CellInsight™ Personal Cell Imager Cytometer (Thermo Scientific) is a high content platform designed for quantitative imaging; it allows automated fluorescence analysis and quantification of proliferative cells. Cellinsight automatically scans 96 well plates and takes images from 15 different fields in each well. On the basis these images the platform can quantify total number of cell and total number of proliferative cells per well.

3.2.10 Statistical analysis for cell proliferation

Description of the data:

1. Transfection experiments were done in five biological replicates. Each biological replicate has three treatments, siRNA1, siRNA2 and Stealth siRNA™ negative universal control, leading to five biological replicates per treatment. For proliferation assays 10 wells of a Nunc-Immuno™ MicroWell™ 96 well solid

plate, containing CTX0E16 or CTX0E03 cells, were transfected for each siRNA treatment, leading to 10 technical replicates for each biological replicates for each cell line.

2. The effect of *TCF4* knockdown on the proliferation of CTX0E16 and CTX0E03 cells was analysed by ki67 inmunostaining. Cell counts for proliferating cells (ki67 positives) and total number of cells per well for each treatment in the five biological replicates (five plates) were obtained from Cellinsight. These values were used in the statistical analyses.

The objective of the statistical analyses was to identify the differences in cell proliferation between cells treated with control siRNA and *TCF4* siRNA conditions. To this end I used a Poisson linear mixed model in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.). This model compares the rate between treatments, taking into account the sample size, including cell counts for proliferating cell and total number of cells per treatment (siRNA1, siRNA2 and negative control siRNA) in the five biological replicates. The analysis considers the systematic variation between plates with a fixed effect per treatment and random effect per plate. Differences of $P < 0.05$ were considered significant.

The exact command using R was:

glmer(count_proliferating_cells ~ offset(log(count_total_cells)) + Treatment + (1|Plate) ,data=data.m2 , family='poisson').

3.3 Results

3.3.1 Transfection efficiency of siRNA.

To determine transfection efficiency, cells transfected with the BLOCK-iT Alexa Fluor® Red Fluorescent Oligo were checked 24 hours post-transfection by fluorescence microscopy. Transfection efficiency was very high with >80% of cells presenting red fluorescence 24 hours post-transfection (Figure 3.1).

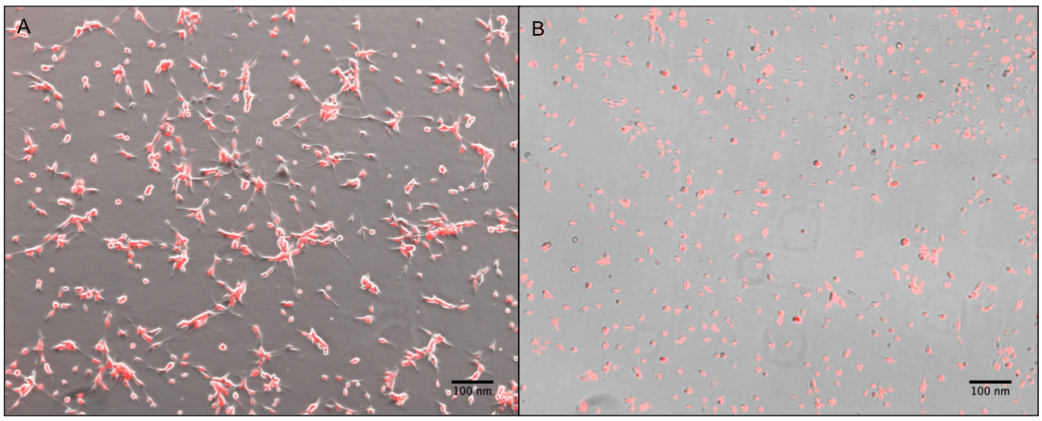


Figure 3.1 Assay of transfection efficiency using the N-TER™ Nanoparticle siRNA Transfection System from Sigma.

Transfection efficiency was measured by the uptake of BLOCK-iT Alexa Fluor red Fluorescent Oligo into neuronal progenitor cells CTX0E03 (A) and CTX0E16 (B). Fluorescence was checked 24 hours post-transfection. Images were taken using the inverted microscope Olympus Ix70 using AxioVision 4.8 Imaging software (Zeiss). Scale bar 100 nm.

3.3.2 Knockdown of *TCF4* mRNA in CTX0E16 and CTX0E03 cells.

Confirmation of *TCF4* mRNA knockdown, relative to the negative control siRNA, was assessed by quantitative qPCR. The qPCR values for *TCF4* mRNA expression level were normalized against the geometric average of three internal control genes (*RPLI3A*, *GADPH* and *HPRT1)*. Table 3.9 and Table 3.10 show the normalized data for CTX0E16 and CTX0E03 cell lines, used in the statistical analyses.

Table 3.9 Normalized *TCF4* expression values obtained in CTX0E16 cell line. These values were used in statistical analysis applying a linear mixed model.

Treatment	Normalized values for TCF4			Biological replicate
	qPCR 1	qPCR 2	qPCR 3	
siRNA1	0.8892	0.8812	0.7748	1
siRNA1	1.0809	0.9022	1.3119	2
siRNA1	0.6614	0.4276	0.5016	3
siRNA1	0.2254	0.1397	0.3057	4
siRNA1	0.2545	0.0826	0.3571	5
siRNA2		0.8365	0.7772	1
siRNA2	0.7370	0.6437	0.7138	2
siRNA2	0.7008	0.4941	0.4140	3
siRNA2	0.2094	0.1693	0.3433	4
siRNA2	0.3859	0.3864	1.2033	5
control	0.8080	0.9786	0.7820	1
control	1.2124	1.0225	0.9007	2
control	0.5042	0.3978	0.3929	3
control	0.7143	0.7145	0.6797	4
control	0.3202	0.4193	0.6579	5

Table 3.10 Normalized *TCF4* expression values obtained for CTX0E03 cell line. These values were used in statistical analysis applying a linear mixed model.

Treatment	Normalized values for TCF4			Biological replicate
	qPCR 1	qPCR 2	qPCR 3	
siRNA1	0.8031	0.6791	0.6266	1
siRNA1	1.0834	0.9797	1.2306	2
siRNA1	0.6731	0.4935	0.6576	3
siRNA1	0.1597	0.1219	0.0949	4
siRNA1	0.3400	0.2880	0.3295	5
siRNA2	0.8907	0.8487	0.9828	1
siRNA2	1.0506	1.0723	0.8381	2
siRNA2	0.5474	0.5629	0.5394	3
siRNA2	0.1881	0.1755	0.2088	4
siRNA2	0.2385	0.2526	0.3655	5
control	0.6591	0.5180	0.8369	1
control	1.2239	1.5834	1.5326	2
control	0.5677	0.5152	0.7036	3
control	1.2455	0.3860	0.6460	4
control	0.2575	0.2206	0.2576	5

Statistical analyses were performed applying a linear mixed model in order to identify expression changes in *TCF4* mRNA level between control siRNA and *TCF4* siRNA conditions. The results obtained from the statistical analysis, for CTX0E16 and CTX0E03 cell lines, are show in Table 3.11 and Table 3.12.

Table 3.11 Statistical results for *TCF4* expression in the CTX0E16 cell line.

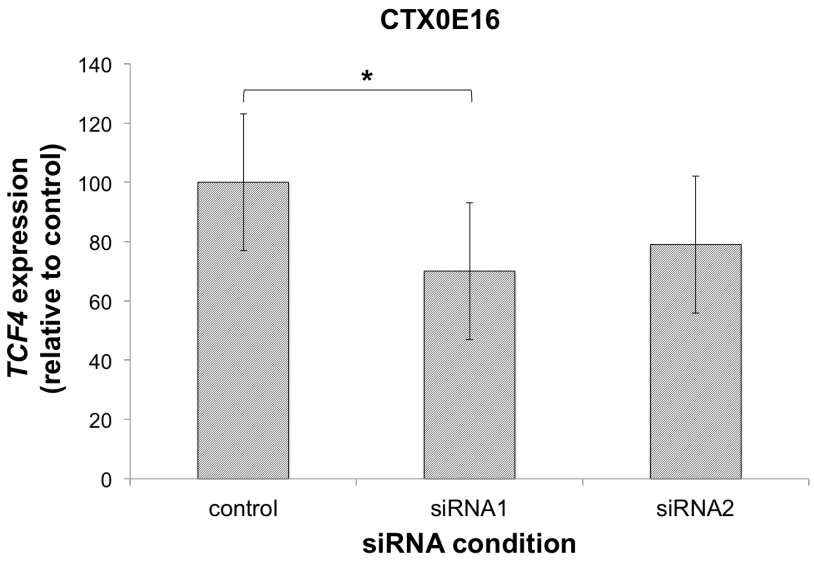
Treatment	Estimate	SE	Pr(>|t|)	Mean expression	mean-1SE	mean+1SE	Fold change (%)
control	-0.4265	0.2310		0.6528	0.5181	0.8224	100
siRNA1	-0.7827	0.2310	0.0347 *	0.4572	0.3629	0.5760	70.0333
siRNA2	-0.6564	0.2332	0.1731	0.5187	0.4108	0.6550	79.4613

Table 3.12 Statistical results for *TCF4* expression in the CTX0E03 cell line.

Treatment	Estimate	SE	Pr(>|t|)	Mean expression	mean-1SE	mean+1SE	Fold change (%)
Control	-0.4759	0.3024		0.6213	0.4592	0.8407	100
siRNA1	-0.8037	0.3024	0.0261 *	0.4477	0.3308	0.6057	72.0507
siRNA2	-0.7295	0.3024	0.0812	0.4822	0.3563	0.6524	77.6002

TCF4 expression was represented relative to the negative control for both cell lines in Figure 3.2. The Figure 3.2 show reduced *TCF4* expression on CTX0E16 cell by 30% and 21% respectively relative to negative control siRNA, being only significant (*P* = 0.0347) for siRNA1. Similarly, a significant reduction of 28% in *TCF4* expression was observed in CTX0E03 cells transfected with siRNA1 (*P* = 0.0261). siRNA2 reduced *TCF4* expression in CTX0E03 cells by 22.4%, although, this was not statistically significant (*P* = 0.0812).

A



B

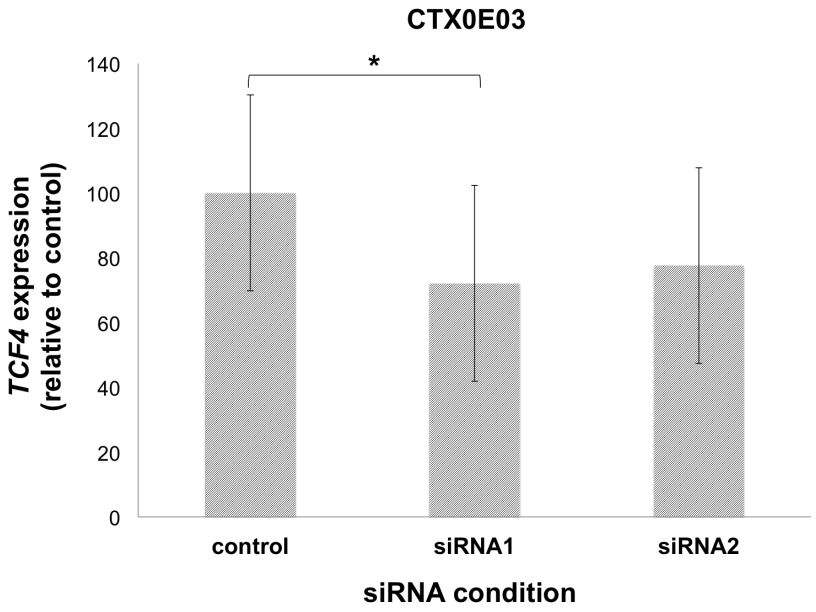


Figure 3.2 Expression of *TCF4* mRNA after RNAi transfection

QPCR result showing *TCF4* mRNA expression 90 hours post-transfection with two independent *TCF4* siRNAs (siRNA1 and siRNA2). Data are normalized against the geometric average of 3 housekeeper genes and plotted relative to the negative control siRNA. *TCF4* mRNA was reduced by 30 % and 21% in association with siRNA1 (P = 0.0347) and siRNA 2 (P = 0.1731) respectively for CTX0E16 cells (A). A reduction of *TCF4* mRNA expression of 28% and 22.4% in association with siRNA1 (P = 0.0261) and siRNA 2 (P = 0.0812) respectively was observed on CTX0E03 (B). Error bars represent SEM derived from five separate transfection experiments. Asterisks shows where the expression of *TCF4* is significantly different from control (P < 0.05).

3.3.3 Effect of *TCF4* knockdown on cell proliferation.

The effect of *TCF4* knockdown on the proliferation of CTX0E16/02 and CTX0E03 cells was analysed by ki67 inmunostaining. Background staining for ki67 was removed by quantifying positive ki67 cells from a negative control (Figure 3.3).

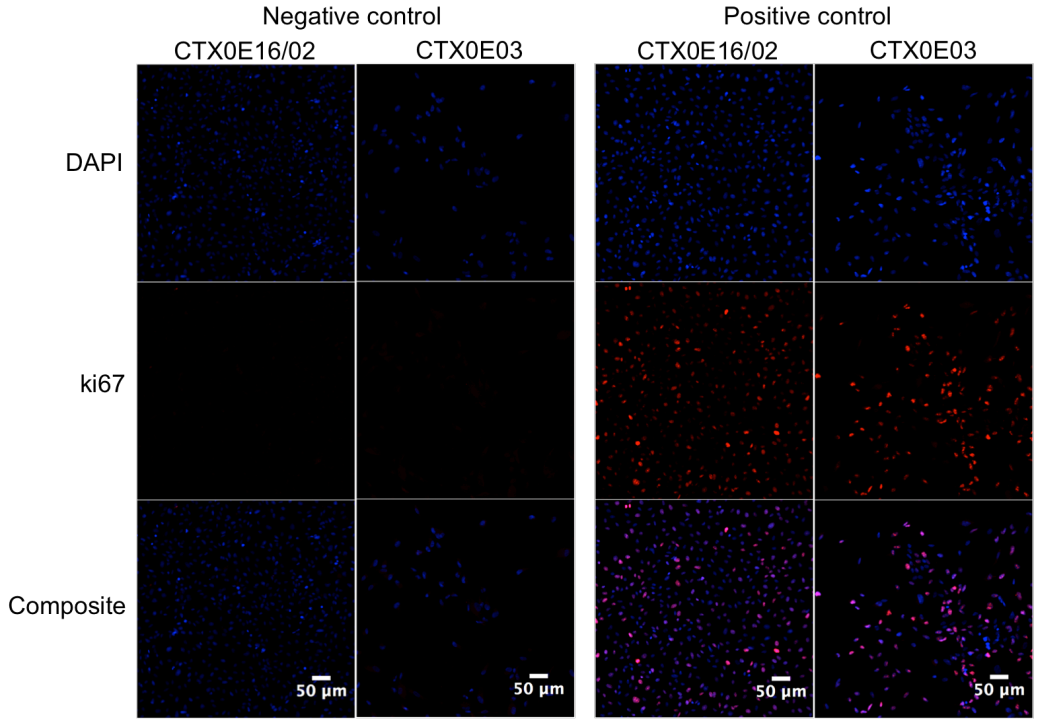


Figure 3.3 Examination of ki67 inmunostaining.

90 hours after transfection cell proliferation was evaluated by ki67 inmunostaining assay. For all conditions a negative control, which was only stained for secondary antibody and DAPI, was used to eliminate background (false positive staining for ki67). Positive ki67 cells were quantified using Cell Insight. Images were taken using Cell Insight (Thermo). Scale bar 50 µm.

Cell counts for proliferating cells (ki67 positives) and total number of cells per well for each treatment in the five biological replicates (five plates) were used in the statistical analyses. Because the number of cells observed per well in each treatment and in the five biological replicates was variable, I decided to analyse my data using a mixed-model to account for this systematic variation. In addition, I used a Poisson linear mixed-model in order to account explicitly for the total number of cells in the well when comparing the fraction of proliferating cells. This helps avoid potentially misleading results due to very different number of cells across treatments or biological replicates. For instance, simply using the fraction of proliferating cells but ignoring the total number of cells would deem a ratio of 50% equal from a well with 100 or 10000 total cells, when in the second case there is much more certainty that the ratio is 50%. The analysis considers treatments as fixed effects and aims to account for the systematic variation between by modelling biological replicates as random effects. Below, I will describe in detail the process used to summarise and filter the data in order to perform statistical analyses as well as checks I did to understand the statistical results and their robustness.

Data analysis for CTX0E16

1. Firstly, I explored the variability in total number of cells per biological replicate and well (Figure 3.4, top). Plate 4 had very few cells for both siRNA and plate 5 had more cells than any other plate for all the three treatments. I also examined the fraction of proliferative cells (Figure 3.4, bottom); plate 4 showed the greatest variability in the raw proliferation ratio.

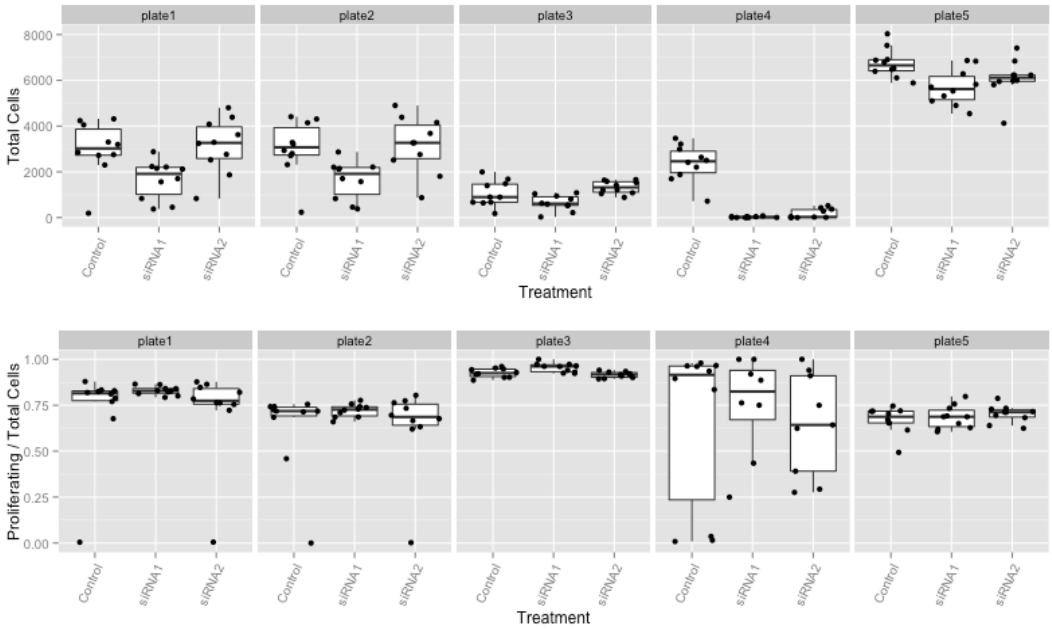


Figure 3.4 Total number of cells for each treatment per plate (top) and fraction of proliferative cells for each treatment per plate (bottom), for CTX0E16 cell line.

2. Then I examined the relationship between total number of cells and number of proliferating cells (Figure 3.5). This confirmed previous observations for plate 4 and showed that some wells have a very low number of proliferating cells, which coincide with wells with low cell counts. It also showed that plate 5 has more proliferating cells than any other plate (all wells are on the top right corner).

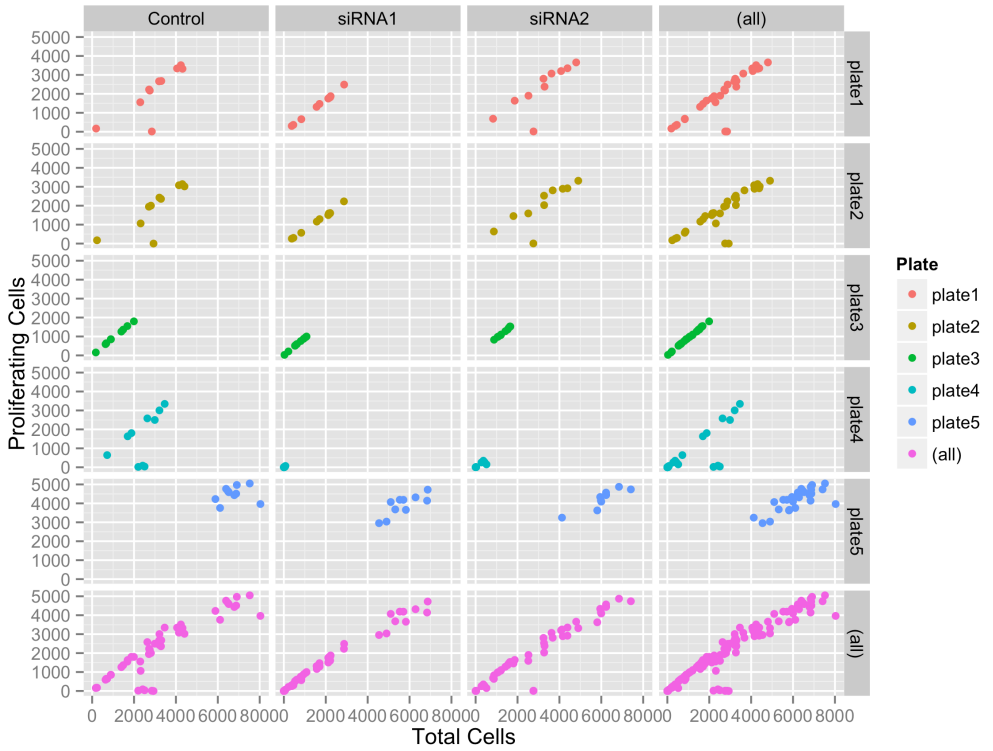


Figure 3.5 Relationship between total number of cells and proliferating cells for each treatment per plate, observed on CTX0E16 cell line

3. Based on observations from point 2, I decided to remove from the analysis wells with less than 100 cells and to remove plate 4 due to the low cell counts observed in the *TCF4* siRNA treatments. This leads to the following dataset: Figure 3.6.

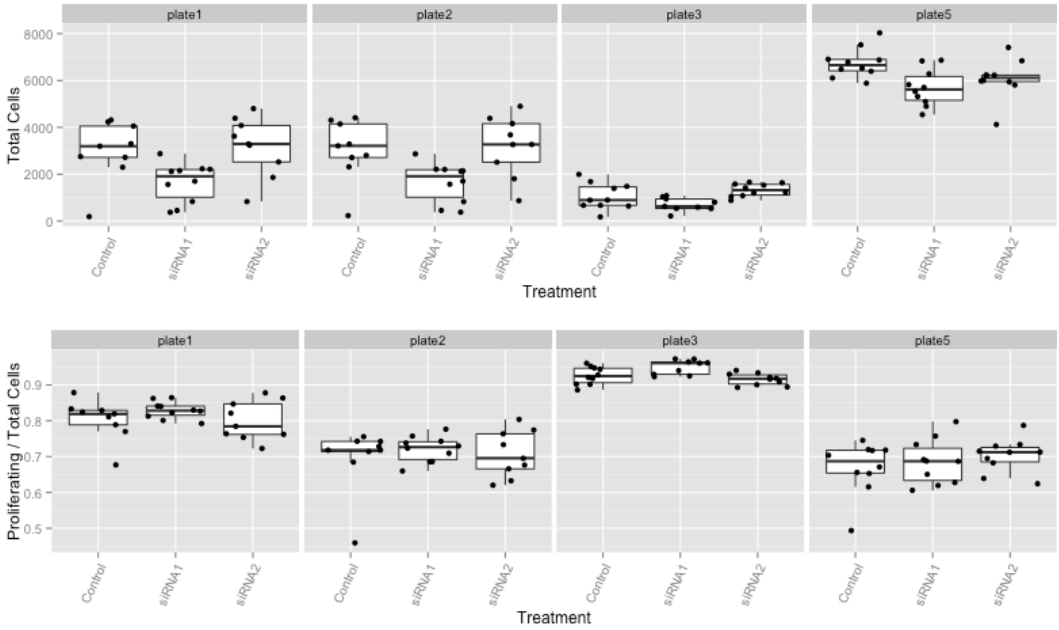


Figure 3.6 Total number of cells for each treatment per plate (top) and fraction of proliferative cells (bottom). These data excluded plate 4 and wells with less than 100 cells. Data obtained on CTX0E16 cell line.

4. In order to take into account the difference in the total number of proliferating cells in different wells I used a Poisson model, which explicitly models this. As explained before, using the ratio of proliferating/total number of cells would ignore differences in total number of cells and regards as equal a case with ratio of 50% from counts 10/20 or 1000/2000 in spite that the second case has a much more certainty that the ratio is 50%. In order to account for the systematic variation across wells and biological replicates I used a generalized linear mixed model.
    a) First I analysed the data removing plate 4 and wells with less than 100 proliferating cells. The statistical results are show in Table 3.13:

Table 3.13 Statistical results for proliferating CTX0E16 cells with plate 4 removed and removed wells with less than 100 proliferating cells.

Treatment	Estimate	SE	Pr(>|z|)	Mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.2714	0.0597		0.7623	0.7181	0.8092	100
siRNA1	-0.2384	0.0598	2.96E-11***	0.7879	0.7422	0.8364	103.3582
siRNA2	-0.2527	0.0597	3.88E-05***	0.7767	0.7317	0.8245	101.8937

b) Finally, I analysed the data removing plate 4 and wells with less than 100 proliferating cells, and removing plate 5 to test if the significance of treatment 'siRNA2' is driven by this plate with unusual high levels of cell count and proliferation: This is the final data to be used for discussion (Table 3.14).

Table 3.14 Final data showing the effect of TCF4 knockdown on CTX0E16 cell proliferation. Statistical results were obtained by removing plate 4 and 5 as well as wells with less than 100 proliferating cells.

Treatment	Estimate	SE	Pr(>|z|)	Mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.2207	0.0631		0.802	0.7529	0.8542	100
siRNA1	-0.1788	0.0632	4.53E-09***	0.8363	0.785	0.8908	104.2801
siRNA2	-0.2272	0.0631	0.2942	0.7968	0.748	0.8487	99.3561

Data analysis for CTX0E03

I repeated the same process used on the CTX0E16 cell line:

1. First I examined the variability in total number of cells per biological replicate and wells (Figure 3.7, top), plate 1 and 5 showed a low number of cells in all the wells and treatments. Then I looked at the fraction of proliferative cells (Figure 3.7, bottom), and observed a general variability in the fraction of proliferative cell across biological replicates. It could be misleading not to account for total number of cells since this varies a lot among wells within and across biological replicates.

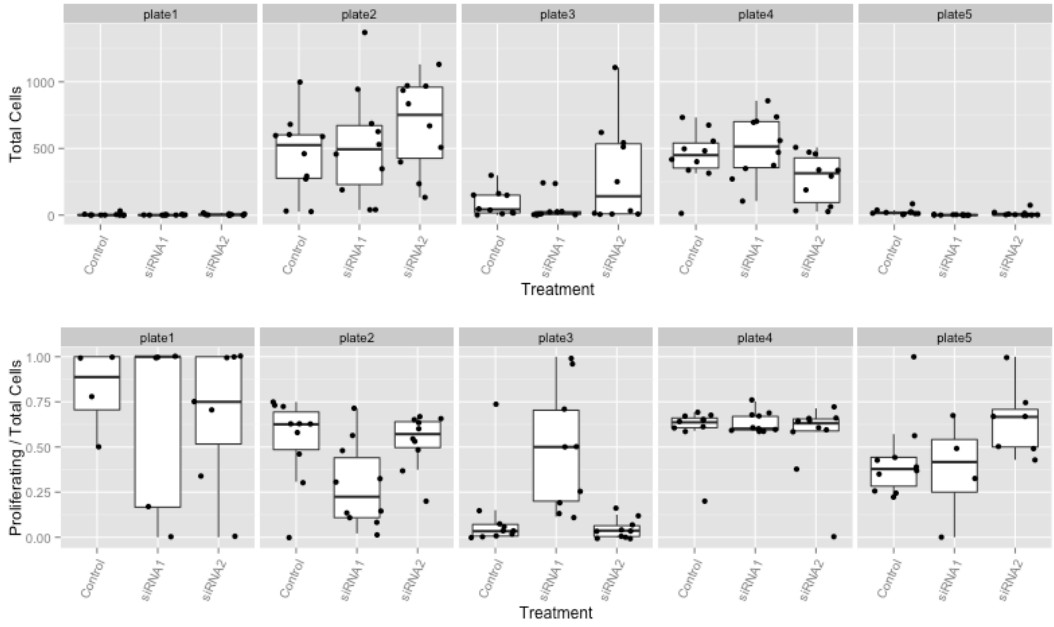


Figure 3.7 Total number of cells for each treatments per plate (top) and fraction of proliferative cells for each treatment per plate (bottom), obtained in the CTX0E03 cell line.

2. Then I examined the relationship between total number of cells and number of proliferating cells (Figure 3.8). A consistent linear relationship between the number of proliferating cells and the total number of cells was observed. This is expected if the cells proliferate at a *constant* rate across biological replicates. However, there were some wells with very low number of proliferating cells. This was also the case for wells with low number of total cells for CTX0E16 cell line.

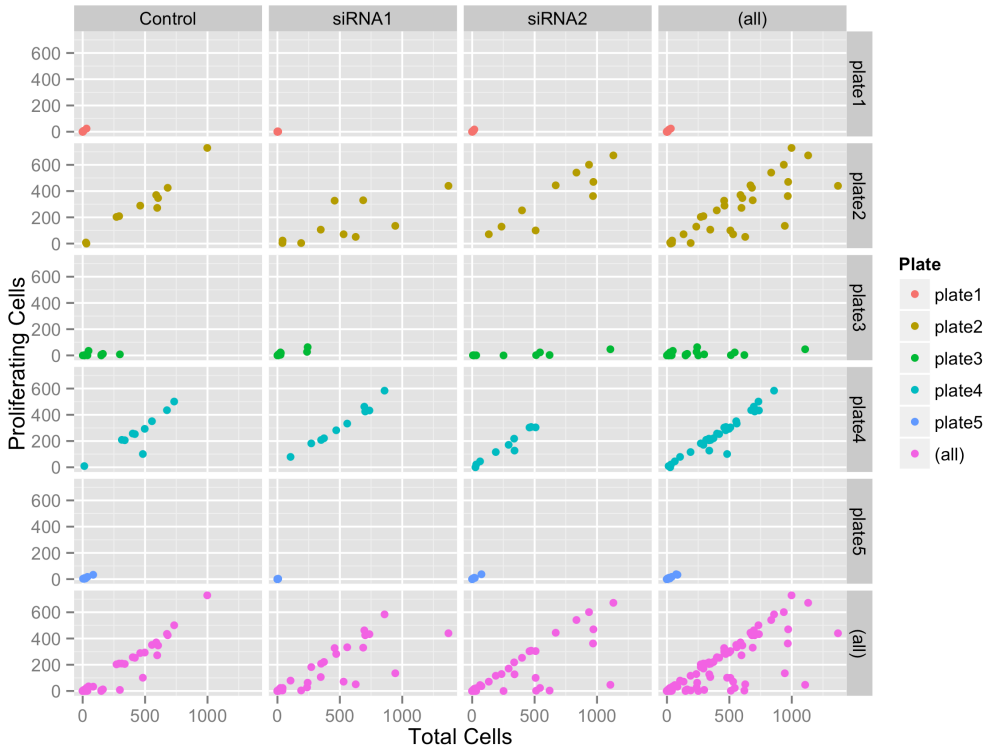


Figure 3.8 Relationship between total number of cells and proliferating cells for each treatment per plate, observed on CTX0E03 cell line.

3. Based on observations from point 2, I decided to remove the wells with less than 100 total cells, since they seemed to behave differently from the rest. This removed completely plates 1 and 5 and some wells from the remaining plates. This lead to the following dataset: Figure 3.9.

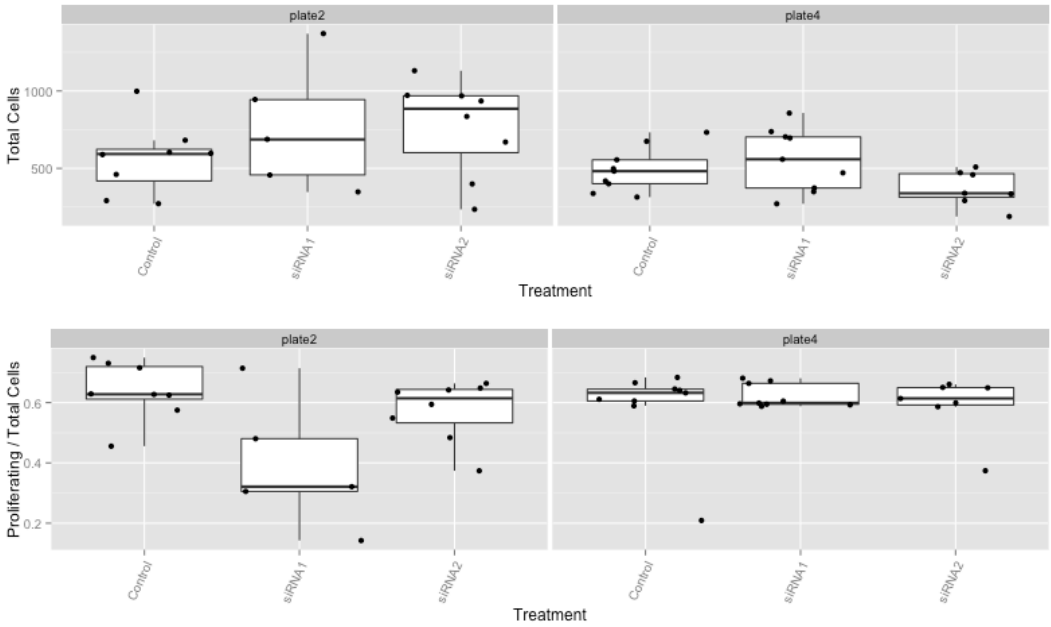


Figure 3.9 Total number of cells for each treatment per plate (top) and fraction of proliferative cells (bottom). These data excluded plates 1 and 5. Data obtained on the CTX0E03 cell line.

4. In order to account for the systematic variation across wells and the two remaining biological replicates I used a generalized Poisson linear mixed model.
    a) First I analysed the filtered data, removing wells with less than 100 proliferating cells; the statistical results are show in Table 3.15:

Table 3.15 Statistical result for proliferating CTX0E03 cells, with wells with less than 100 proliferating cells removed.

Treatment	Estimate	SE	Pr(>|z|)	Mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.4933	0.0563		0.6106	0.5772	0.6460	100
siRNA1	-0.6940	0.0567	3.25E-23***	0.4996	0.47219	0.5287	81.8202
siRNA2	-0.5263	0.0565	0.0964	0.5908	0.5583	0.6251	96.7527

b) Finally, in order to compare I re-analysed the data without filtering, the statistical results are show in Table 3.16:

Table 3.16 Statistical result for proliferating CTX0E03 cells, without filtering.

Treatment	Estimate	SE	Pr(>|z|)	mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.9388	0.3980		0.3911	0.2627	0.5823	100
siRNA1	-1.2065	0.3981	2.03E-42***	0.2992	0.2010	0.4456	76.5127
siRNA2	-1.0020	0.3980	0.0011**	0.3671	0.2466	0.5466	93.8703

The comparison of both statistical results suggests that both results are consistent. Without filtering the data siRNA2 would be declared significant ($P < 0.05$). Given that this result is only seen with the data I considered not reliable was included, I did not consider the results obtained by analysing all the data robust. For the CTX0E03 cell line, the final data used for discussion was therefore the filtered data for >100 cells.

Summary of proliferation results for CTX0E16 and CTX0E03 cell line:

The fraction CTX0E16 proliferative cells increased when *TCF4* was knocked down with siRNA1 and siRNA2, relative to the negative control siRNA condition. Table 3.14 shows an increase in cell proliferation occurring in association with siRNA1 and siRNA2, although this was only significant for siRNA1 ($P = 4.53 \times 10^{-9}$, fold change 104.3%). In contrast, when CTX0E03 were transfected with *TCF4* siRNAs a decrease in cell proliferation was observed in association with siRNA1 and siRNA2, which was again only significant for siRNA1 ($P = 3.25 \times 10^{-23}$, fold change 81.8%), Table 3.15.

3.4 Discussion

The main aim of the experiments described in this Chapter was to assess the effect of downregulating endogenous *TCF4* on the proliferation of two karyotypically normal neural progenitor cell lines derived from human foetal brain, CTX0E16 and CTX0E03, and determine if the direction of the effect differs between the cell lines. To this end I used RNAi followed by cell proliferation assays.

TCF4 knockdown altered the proliferation of both cell lines, with one cell line showing decreased proliferation and the other showing increased proliferation. Interestingly, the effect on proliferation and the direction of the effect is consistent with the gene expression data obtained by Dr Nick Bray's group. Both lines of evidence suggest a cell-specific effect of TCF4 knockdown. I have two possible explanations for these results. The first one is that the two cell lines are expressing different *TCF4* transcripts that are being knocked down. A previous study showed that *TCF4* gene has numerous exons and produces several transcripts resulting in a differential subcellular distribution and transactivational capacity of TCF4 isoforms (Sepp et al., 2011). However, RT-PCR data obtained by James Docker, a former MSc student in Dr Bray's Lab, found no differences in the *TCF4* transcripts being expressed by both cell lines (Figure 3.10).

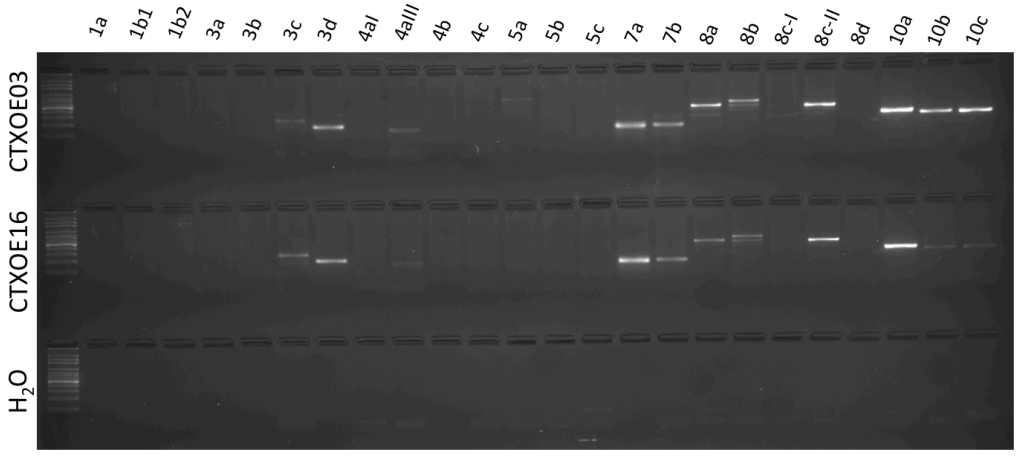


Figure 3.10 Expression of alterative *TCF4* transcript in CTX0E16 and CTX0E03 neuronal progenitor cell lines. RT-PCR analysis of *TCF4* transcripts with different 5' exons.

A second possible explanation is that both cell lines are expressing different interactors of TCF4 and this may cause the opposing effects. As previously described, TCF4 can act as a transcriptional activator or repressor depending on the heterodimers formed with other bHLH proteins (Murre, 2005; Sweatt, 2013), and TCF4 subcellular localization is also dependant on its interactors (Sepp et al., 2011). Furthermore, heterodimers with tissue specific bHLH proteins are essential in regulating the function of bHLH proteins and determining which specific DNA sites will be recognized (Goldfarb et al., 1996). Consistently, the micro-array data obtained by Dr Nick Bray show that both cell lines expressed a distinctive group of bHLH genes (assessed in the control cells). 44 bHLH genes are commonly expressed in both cell lines of the total of 58. 3 and 5 bHLH genes are unique for CTX0E03 and CTX0E16 respectively. Unique bHLH genes for CTX0E03 cell line are, *BHLHB3*, *HES2* and *SOHLH2*; and for CTX0E16 cell line, *ARNT*, *ATOH8*, *MYCN*, *OLIG2* and *SREBF2*. Additionally, 10 out of the 37 bHLH genes expressed in both cell lines show significant differences in expression between them (Bonferroni corrected $P < 0.05$, for 45 probes tested). Of these, ID1 is an interesting candidate to explain the opposite behavior of the two cell lines. Parrinello et al (Parrinello et al., 2001) provides evidence suggesting that the effect of ID1 in mammary epithelial cell proliferation occurs through its interaction with TCF4. They observed a reduction of ID1 stimulated proliferation when TCF4 was overexpressed; suggesting that higher expression of ID1 is associated with higher proliferation. Interestingly, in the present study the CTX0E16 cell line showed higher proliferation than CTX0E03 in the control condition and this correlated with ID1 expression. ID1 expression is on average 16% higher for CTX0E16 compared with CTX0E03 ($P = 0.0001$, Bonferroni corrected $P = 0.0045$, for 45 probes tested). This suggests that interaction between TCF4 and ID1 might account for the differences in proliferation observed between both cell lines. However, more studies are necessary to confirm this observation and the mechanism for these effects.

In addition, preceding studies revealed that TCF4 regulates the expression of two genes involved in cell cycling and proliferation, namely $p57^{kip2}$ and $p21^{cip1}$ (Herbst et al., 2009;

Schmidt-Edelkraut et al., 2014). These studies were conducted in mouse neuronal progenitors and colorectal cancer cells lines for $p57^{kip2}$ and $p21^{cip1}$, respectively. As ID1, $p57^{kip2}$ and $p21^{cip1}$ are candidates to explain the mechanism by which TCF4 effects cell proliferation, it would be interesting to investigate the expression of these in the two cell lines.

Several studies have focused on the role of cell proliferation in the pathogenesis of schizophrenia. Reif et al (Reif et al., 2006) report reduced ki67 staining in post-mortem brain from patients with schizophrenia, but not depression, compared with that from controls. Additionally, Wang et al, (Wang et al., 2010) demonstrated altered cellular proliferation and growth pathways in fibroblast obtained from schizophrenia patients and recent studies in olfactory neurosphere-derived cells from schizophrenia patients showed an association between altered cell cycle and schizophrenia (Fan et al., 2012). Patients with schizophrenia showed increased cell proliferation associated with altered dynamics of expression of cyclin proteins. Furthermore, there is evidence showing changes in cell proliferation mediated by drugs used to treat schizophrenia, such as valproic acid, olanzapine and others (Dodurga et al., 2014; Kimoto et al., 2011; Kurihara et al., 2014; Niu et al., 2010; Schwarz et al., 2008).

Here, knockdown of *TCF4* in two neuronal progenitor cell lines provides evidence for its involvement in the regulation of cell proliferation. These physiological data strongly complement the gene expression data previously obtained by Dr Bray's Lab and accord well with the known biology of TCF4 and multiple lines of evidence associating altered cell proliferation with schizophrenia pathogenesis. Altogether, these findings suggest that allelic variation on the *TFC4* loci might affect proliferation of neural cells and in this way contributes to the pathogenesis of schizophrenia. The differential effect on the two cell lines highlights how cellular consequences of altered gene function can depend on cell type, potentially, among other factors, due to interacting genes showing cell type differential expression. ID1, $p57^{kip2}$ and $p21^{cip1}$ constitute potential candidates to further elucidate this cell type specific effect.

Limitations of the study:

The present study is the first one to provide experimental evidence for the biological effect of TCF4 knockdown in neural progenitor cell lines derived from human foetal brain. However, in order to understand the cell specific trasncriptional effect of TCF4 knockdown follow up studies in mature neuronal cells will be needed.

Additionally, the present study fail to show the relevance of the risk variant effect for schizophrenia. So far studies have been unsuccessful in finding association between the risk variants and TCF4 expression in adult brain. This may reflect outcomes for other schizophrenia risk SNP exerting its effect only in foetal brain such in ZNF804a (Hill and Bray, 2012) and VRK2 (see Chapter 5).

Chapter 4. Allelic expression of *VRK2* in the human brain

4.1 Introduction

VRK2 is a member of the vaccinia-related kinase (VRK) family of serine/threonine kinases and is located at chromosome 2p16.1. *VRK2* gene is widely expressed in human tissues, especially those that are highly proliferative (Nezu et al., 1997). Two splice variants generate isoforms A and B of 508 and 397 amino acids, respectively (Figure 4.1). The specific subcellular location of these isoforms is dependant of cell type (Blanco et al., 2006).

The single nucleotide polymorphism (SNP) rs2312147 is located at chromosome 2p15.1 within the first intron of the long splicing isoform of *VRK2* and about 50kb upstream of the short isoform. The C-allele of this variant has been associated with increased risk to schizophrenia in multiple GWAS. Stefansson et al. (Stefansson et al., 2009) initially reported the association at rs2312147 with a p-value not reaching genome-wide significance (odd ratio OR = 1.09, $P = 3.2 \times 10^{-7}$) but a follow up study provided robust statistical evidence of association with rs2312147[C] (odd ratio OR = 1.10, $P = 5.4 \times 10^{-10}$) (Steinberg et al., 2011). This association has gained additional support in a recent meta-analysis in Asian populations (Li et al., 2012) and is supported by replication studies (Irish Schizophrenia Genomics and the Wellcome Trust Case Control, 2012; Li et al., 2012).

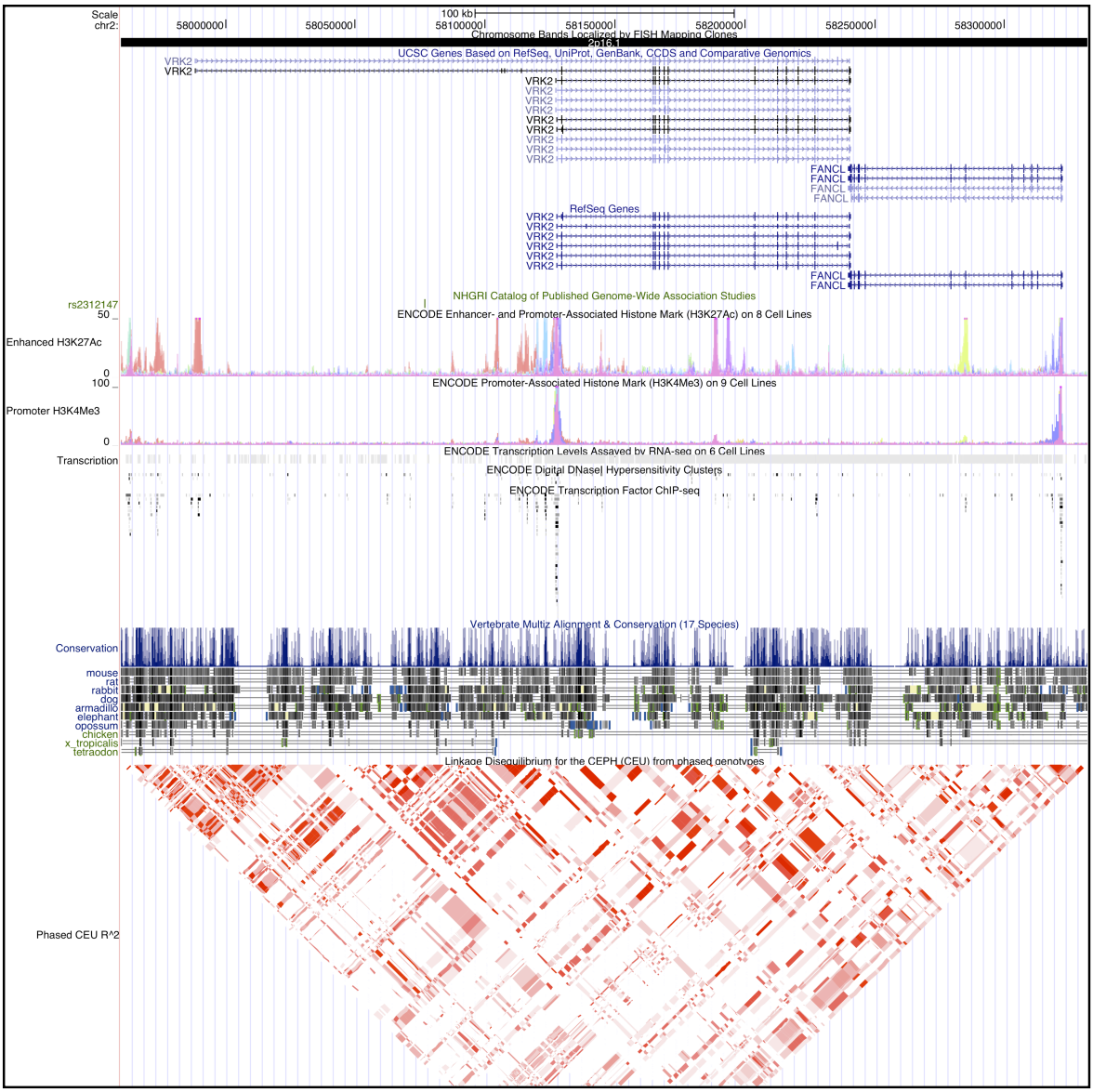


Figure 4.1 Genome browser view of *VRK2*.

Information displayed from top to bottom is: scale, chromosome position, chromosome band, *VRK2* gene models provided by UCSC, Schizophrenia associated SNP, Histone binding determined by Chip-Seq (ENCODE project), promoter activity (ENCODE project), transcription activity (ENCODE project), DNase I hypersensitivity (ENCODE project), transcription factor binding sites determined by Chip-seq (ENCODE project), multiple species conservation and comparisons, linkage disequilibrium in HapMap CEU samples.

Rs2312147 has been associated with brain structure variations in healthy subjects, including total brain and white matter volume, with individuals carrying the risk C-allele tending to have smaller volumes (Li et al., 2012). However, as with most loci associated by GWAS, the molecular mechanisms mediating the statistical association at rs2312147 have not been clearly elucidated. Given its location in noncoding sequence close to the first exon it is plausible that this genetic variant (or another in LD with it) affects the regulation of the expression of the gene.

Differences in gene expression are an important determinant of human phenotypic variability including susceptibility to complex diseases (Bray et al., 2003; Dermitzakis, 2008). For example, it has been shown that schizophrenia susceptibility alleles are en masse enriched for alleles that affect gene expression in brain (Richards et al., 2012). A gene's expression is regulated by both *trans* and *cis*-acting influences. *Trans*-acting regulation operates on both chromosomal genes copies through a remote source (e.g., transcription factors), generally affecting expression of each allele to the same extent. *Cis*-acting influences reside in the same DNA molecule as the regulated gene (Figure 4.2) and are allele-specific. Importantly, both regulatory mechanisms are affected by genetic variation and epigenetic marks (Bell et al., 2011; Dimas et al., 2008; Dixon et al., 2007b; Stranger et al., 2007).

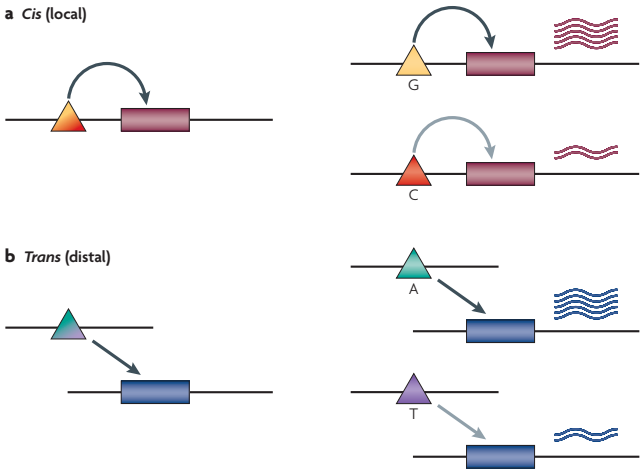


Figure 4.2 *Cis* and *trans* regulatory variations. a) A *cis*-regulatory variant gives more mRNA from the gene copy carrying the G allele than the C allele. b) *Trans* regulatory variant can have different affinity for a target gene leading to different levels of RNA transcription (taken from Cheung and Spielman, (Cheung and Spielman, 2009)).

Rs2312147 could regulate *VRK2*'s transcription and therefore affect its mRNA levels in *cis*. This can be determined using an eQTL (expression quantitative trait loci) approach, where measurements of total gene expression are correlated with genotype in order to estimate the functional contribution of regulatory variants (Nica and Dermitzakis, 2008). However, in a previous study, Li et al (Li et al., 2012) did not found an association of rs2312147 with gene expression in blood samples.

It is important to consider that measures of total expression are complicated and fail to distinguish between *cis*-acting and *trans*-acting influences (Stamatoyannopoulos, 2004). The eQTL approach is prone to all types of confounds reducing the power to detect *cis* effects of SNPs, particularly in post-mortem brain. Some of these confounding factors are: differences in tissue and mRNA quality, cellular composition of samples, environmental exposure such as drugs or pre-agonal state and trans-effects of other gene loci. Furthermore, a recent study (Almlof et al., 2012) showed that eQTL

approaches required up to 8 times more samples than allelic expression measures to detect the effects of the same *cis*-regulatory variants. This study was performed using lymphoblast where the tissue variables are largely controlled.

Analysis of allelic expression appears as an alternative and powerful method to indirectly detect influences of specific haplotypes on a gene's expression (Yan et al., 2002). It has been successfully used to detect differences in allelic expression of other schizophrenia susceptibility genes in adult and foetal brain (Buonocore et al., 2010; Hill and Bray, 2012), supporting its ability to generate new insight into the fundamental biology of neuropsychiatric phenotypes.

Assays of allelic expression make use of a SNP localised in an exonic region of the gene to quantify and discriminate RNA molecules transcribed from each chromosome in heterozygous subjects (Figure 4.3). This method focuses on allelic variation within a subject so each allele serves as an internal control while controlling for potential trans-acting influences, like tissue preparation, RNA quality or environmental influences between others (Bray et al., 2003). A significant departure from the expected 1:1 allele ratio in any assayed sample denotes the presence of *cis*-acting regulatory variation (Bray et al., 2003; Yan et al., 2002).

Interestingly, previous studies have revealed that the effects of *cis*-regulatory variations can differ between tissues, cell types and even developmental stages (Buonocore et al., 2010; Cowles et al., 2002; Dimas et al., 2009; Hill and Bray, 2012; Zhang et al., 2009), making analyses of allelic expression an attractive tool to determine when and where a risk variant affects the expression of a gene.

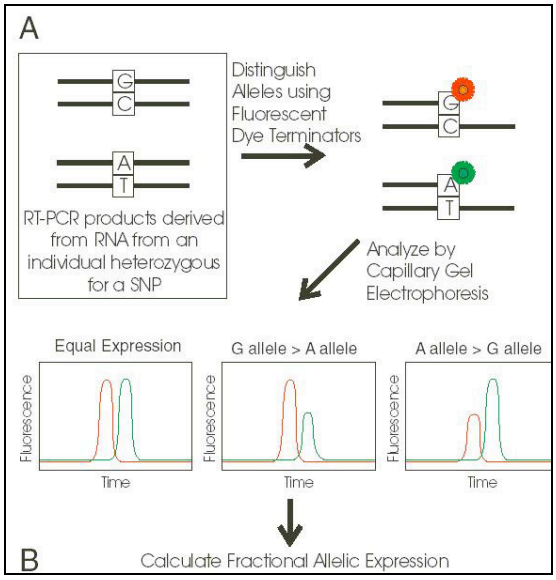


Figure 4.3 Principles of relative allelic expression.

In the absence of *cis*-acting regulatory variation, each copy of a gene (both alleles) will be equally expressed. Conversely, if the subject is heterozygous for a *cis*-acting variant affecting the expression of the gene, the mRNA from each copy is expressed differentially (Reproduced from http://www.icarda.org/generationcp/cp-1-allele.htm).

In this study, the effect of the schizophrenia associated polymorphism rs2312147 on *VRK2*'s allelic expression imbalance was tested by comparing homozygotes and heterozygotes for this SNP. If rs2312147 was associated with altered gene expression a greater relative difference in the expression of each gene copy would be observed in heterozygous individuals compared with homozygotes (Bray and O'Donovan, 2006).

4.1.1 The aim of the present study

The aim of experiments presented on this chapter was to test two hypotheses: a) that *VRK2*'s expression is subject to variable *cis*-regulatory influences in the human brain, and b) that rs2312147 is associated with altered *cis*-regulation of *VRK2* in the human brain. To this end I performed allelic expression assays on dorsolateral prefrontal cortex (DLPFC) in adult human brain and two stages of foetal brain development.

4.2 Methodology

4.2.1 Brain Samples

Adult brain samples

Post-mortem adult human brain tissue was obtained from the Medical Research Council London Neurodegenerative Diseases Brain Bank. Genomic DNA had previously been extracted from 96 of these subjects by Dr N. Bray and was made available for this study. All subjects were free from a psychiatric or neurological diagnosis at the time of death. The allelic expression assay was performed using RNA derived from the dorsolateral prefrontal cortex (DLPFC) of 35 of these subjects who were identified by genotyping (see section 4.2.9) as heterozygous for the expressed SNP rs1051061. RNA was already available from 24 of these subjects within Dr Bray's lab. Tissue from the DLPFC of the remaining 11 heterozygous subjects was requested from the brain bank and RNA extracted (see section 4.2.2). Demographics for the 35 adult subjects assayed for *VRK2* allelic expression are shown in Table 4.1.

Foetal brain samples

Human foetal brain tissue from terminations of pregnancy was provided by the Human Developmental Biology Resource, Institute of Child Health, University College London. Genomic DNA from 40 first trimester and 110 second trimester foetal brain samples had previously been extracted by Dr N. Bray and Dr M. Hill was made available for this study. The allelic expression assay was performed using RNA derived from the whole brain of 10 first trimester foetuses and 38 second trimester foetuses that were identified by genotyping (see section 4.2.9) as heterozygous for the expressed SNP rs1051061. RNA was already available from all samples within Dr Bray's lab. Demographics of the 10 first trimester and 38 second trimester samples assayed for *VRK2* allelic expression are shown in Table 4.1.

Ethical approval for this study was obtained from the Joint South London and Maudsley and the Institute of Psychiatry National Health Service Research Ethics Committee PNM/12/13-102.

Table 4.1 Demographics of subjects assayed for *VRK2* allelic expression in the present study.

Study group	Size (N)	Sex (N)		Age at Death			Post mortem delay (h)	
		M	F	Mean		Range	Mean	Range
Adult Brain DLPFC	35	21	14	70.8 years	(17)	25-102 years	30.2 (15.3)	5 - 72
Foetal Brain								
second trimester	38	17	21	15.3 weeks	(2.9)	13-26 weeks	- -	- -
first trimester	10	-	-	9.8 weeks	(1.9)	7-12 weeks		

N= the total number of subjects used in the study. Sex: F=female, M= male. Numbers in parenthesis represent the standard deviation.

4.2.2 RNA extraction and cDNA synthesis

Total RNA was extracted from frozen brain samples using the Tri-Reagent® Solution (Ambion, Life Technologies, Grand Island, N.Y.). Each human brain sample was stored at -80 °C in Lysing Matrix D tubes (MP Biomedicals). ~100 mg per sample was homogenized with 1 mL of Tri-Reagent® Solution using FastPrep® -24 Instrument (MP Biomedical) for 45 seconds at speed setting of 5. After homogenization, samples were transferred to a clean microcentrifuge tube. Total RNA was extracted following the protocol described in section 3.2.3 (Chapter 3). RNA concentration and purity was quantified using the NanoDrop™ 1000 spectrophotometer.

DNAse treatment of total RNA samples was performed with TURBO DNA-*free*™ (Ambion) (as described in section 3.2.4, Chapter 3). Two RT reactions were performed for each sample. cDNA was synthesized from 1 µg total RNA from each sample using SuperScript™ III reverse transcriptase (Invitrogen) and random decamers (Ambion) in final reaction volume of 20 µL (a detailed protocol can be found in section 3.2.6, Chapter

3). The resulting cDNA samples were diluted 1/7, by adding 120 µL of nuclease-free water (Sigma-Aldrich) to each 20 µL of cDNA, before assaying allelic expression. This concentration is referred to in this thesis as the "standard cDNA concentration". cDNA samples were stored at -20 °C

4.2.3 Primer design

PCR primers used to amplify the region containing the SNPs of interest (Table 4.2) were designed using the Primer3Plus program available online (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). Input DNA sequences were retrieved from the Chip Bioinformatics Tool website (http://snpper.chip.org/), the UCSC genome browser (http://genome.ucsc.edu/) and from the ENSEMBL genome browser (http://www.ensembl.org/index.html).

Table 4.2 Details of primers used for genotyping and in the allelic expression assay.

Polymorphism	Primer	Sequence (5' - 3')
rs1051061	Forward	GATGTACTGGAATATATCATG
	Reverse	CTGGTCTGGATTTTTGTAAC
	Extension	GAATATGTTCATGGTGAT
rs2312147	Forward	CAAGCATCCAGATTTAATCAAAA
	Reverse	TCCAGGCAGTAGTACCAAAGTG
	Extension	TGAAACTGCTTTTCTTTC

4.2.4 Polymerase Chain Reaction (PCR) protocol

PCR reactions were performed in a total reaction volume of 12 µL, containing 6 µL of genomic DNA (8 ng/µL), cDNA or nuclease-free water (Sigma). Nuclease-free water was used in as negative control PCR reaction to check for contaminating templates. FIREPol® DNA Polymerase 5X mix (SolisBioDyne) was used to carry out the PCR reactions. FIREPol® is a highly processive, thermostable DNA polymerase, with an

enhanced stability at room temperature. It has 5'→3' polymerization-dependent exonuclease replacement activity but lacks 3'→ 5' exonuclease activity. Table 4.3 shows details of volumes and concentrations of the reagents used per PCR reaction.

Table 4.3 Reagents used in each PCR reaction.

Reagent	Supplier	Concentration	Volume (µL)
FIREPol® DNA Polymerase mix (0.4 M Tris-HCl, 0.1 M (NH4)2SO4, 0.1% w/v Tween-20, 12.5 mM MgCl2 2.5, 2 mM dNTPs of each – 200 µM dATP, 200 µM dCTP, 200 µM dGTP and 200 µM dTTP)	Solis Biodyne	5x	2.4
Forward primer	Sigma-Aldrich	10 µM	0.4
Reverse primer		10 µM	0.4
Nuclease-free water		-	2.8
DNA template/nuclease-free water		-	6
Total	-	-	12

PCR reactions were performed in 96-well plates in a G-Storm thermal cycler (Somerton Biotechnology Centre, UK), with 35 cycles and annealing temperatures between of 50 °C and 60 °C according to the primers used.

Annealing temperature for each set of primers was optimized by using a thermal cycler block with a gradient of annealing temperatures between 50 °C and 60 °C. PCR amplification was checked by agarose gel electrophoresis and UV visualization. The optimal annealing temperature was defined by the strongest band, which has the right size without non-specific amplification. Table 4.4 shows a typical thermal cycling program.

Table 4.4 Typical PCR cycling conditions.

Step	Temperature (°C)	Time (minutes)	Number of Cycles
Initial denaturation	95	5	1
Denaturation	95	0.4	35
Annealing	50	0.30	
Extension	75	0.4	
Final extension	74	10	1
Cool down	4	5	

4.2.5 Agarose gel electrophoresis

Agarose gels were made at 3% (w/v) agarose (Invitrogen) in 1 X TAE buffer (40 mM Tris-acetic acid, 10 mM EDTA pH 8.0). The agarose was dissolved in 1 X TAE buffer by heating in a microwave. The solution was cooled down before adding ethidium bromide (EtBr) to a final concentration of 0.5 µg/mL.

DNA samples were prepared for electrophoresis by mixing 4 µL of each PCR product with 4 µL of Orange G gel loading buffer before loading into the pre-made wells in the agarose gel. 1 µL of a 100 base pair (bp) DNA ladder (Solis BioDye), containing 12 DNA fragments in a range of 100 – 3000bp, was run alongside the samples.

The gel was placed into an electrophoresis tank and submerged with 1 X TAE and it was run at 120 V for approximately 30 minutes.

EtBr staining of the acid nucleic was visualized using the UV transilluminator system BioSpectrum® Imaging System (UVP, USA) and pictures of the gels were taken using the software Vision Works LS (UVP, USA).

4.2.6 Genotyping

Genomic DNA samples were genotyping for the *VRK2* expressed 'tag' SNP rs1051061 and the schizophrenia risk SNP rs2312147. Genotyping was carried out by a single base primer extension using the SNaPshot Multiplex Kit (Applied Biosystems). The SNaPshot protocol comprises PCR, post-PCR clean up, SNaPshot primer extension, and capillary gel electrophoresis. A detailed description of each of these steps is found in the next section. SNP genotype data was analysed with GeneMarker software v1.91 (SoftGenetics).

4.2.7 Allelic expression assays

Assays of relative allelic expression makes use of a SNP localised in an exonic region of the gene to quantify and discriminate RNA molecules transcribed from each chromosome in heterozygous subjects. The exonic SNP rs1051061 was used to assay *VRK2*. This SNP was in moderate linkage disequilibrium with the schizophrenia risk SNP rs2312147 (combined sample D' = 0.82; r^2 = 0.20). Because the LD was not complete (i.e. $r^2 < 1$), the effect of the risk SNP rs2312147 genotype on *VRK2* allelic expression could be tested by further separating our assayed rs1051061 heterozygotes into two groups, those that were homozygous and those that were heterozygous for the schizophrenia risk SNP rs2312147.

First, individuals heterozygous for the exonic SNP rs1051061 were identified by genotyping genomic DNA from the 96 adult brain samples and 150 foetal brain samples (first and second trimester). cDNA from the brain RNA of heterozygous subjects was then assayed for VRK2 allelic expression, alongside the corresponding heterozygous genomic DNA samples and negative controls. The assay was carried out by a single base primer extension using the SNaPshot Multiplex Kit (Applied Biosystems). SNaPshot protocol comprises PCR, post-PCR clean up, SNaPshot primer extension, and capillary gel electrophoresis. Figure 4.4 summarise the SnaPshot protocol.

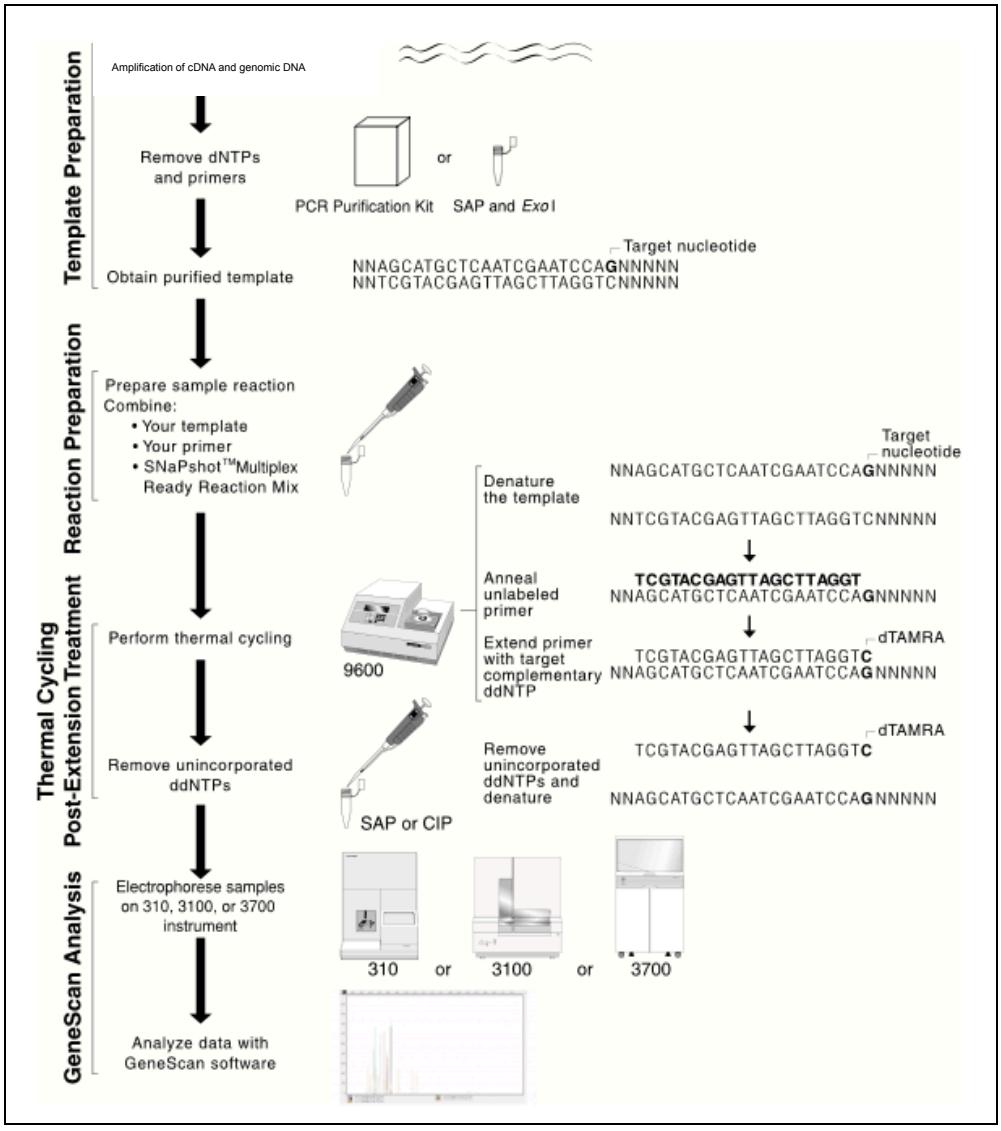


Figure 4.4 Protocol to measure differential allelic expression.

(Taken from ABI PRISM® SNaPshot™ Multiplex Kit Protocol 4323357B, Applied Biosystems)

A detailed explanation of each step can be found below:

PCR and post-PCR cleanup

Sequence containing the exonic SNP rs1051061 was PCR-amplified twice from two RT reactions products from each brain sample, two heterozygous genomic DNA samples and two negative controls. PCR primers were based on a single exon sequence, designed to produce the same amplicon from both genomic DNA and cDNA (Table 4.2).

PCR was carried out in a total reaction volume of 12 µL, containing 6 µL of genomic DNA (8 ng/µL), cDNA or nuclease-free water (Sigma). The PCR protocol is described in detail in section 3.2.8. To confirm amplification and no contamination, 4 µL of each amplified sample was electrophoresed on an agarose gel and EtBr staining of the acid nucleic was visualized by UV transillumination. PCR reactions were then incubated with 1 µL (1U) of rAPid Alkaline Phosphatase (Roche), 0.2 µL (2U) Exo-I (New England Biolab) and 1.8 µL of sterile water (Sigma-Aldrich) for 45 minutes at 37 °C and then for 15 minutes at 85 °C prior to primer extension reactions.

SNaPshot primer extension

Primer extension reactions were performed in a total volume of 10 µL, containing 2 µL treated PCR product, 1,25 µL SNaPshot Multiplex Kit (Applied Biosystems), 5.75 µL nuclease free water (Sigma) and 1 pM extension primer (sequence on Table 4.2). Primer extension thermal cycling conditions are described on Table 4.5. The reaction was carried out in a 96-well plate in a G-storm thermal cycler.

Table 4.5 Thermal cycling protocol for SNaPshot primer extension.

Step	Temperature (°C)	Time (minutes)	Number of Cycles
Initial denaturation	95	2	1
Denaturation	95	0.5	30
Annealing	43	0.5	
Extension	60	0.5	

Capillary electrophoresis of SNaPshot products

Capillary electrophoresis was used to separate DNA fragment fluorescently labelled by SNaPshot primer extension. Following primer extension, aliquots of 2 µL of SNaPshot reaction was combined with 8 µL of Hi-Di formamide (Applied Biosystems) and electrophoresed on an ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) on a 36 cm capillary array with POP-7 ™ Polymer (Applied Biosystems), controlled by data collection software 3.0 (Applied Biosystems).

Peaks heights of allelic specific extended primers were determined using GeneMarker software (SoftGenetics). The allele ratio of the two peaks heights was calculated for each reaction by dividing the peak height of the minor allele by the peak height of the major allele. This ratio was used as an indicator of the relative expression of the two alleles in each individual sample.

It could be assumed that genomic DNA should reflect a perfect 1:1 ratio of the two alleles (Bray et al., 2003). Consequently, for each plate, the average ratio of all genomic DNA was used to correct the ratios obtained from all cDNA and genomic DNA samples for any inequalities in allelic representation specific to each assay. This technique was repeated four times for each sample, giving a total of four corrected ratios for cDNA and genomic DNA per sample.

4.2.8 Statistical analysis

The effect of genotype at schizophrenia risk SNP rs2312147 on *VRK2* allelic expression was tested by comparing corrected cDNA allele ratios from rs2312147 heterozygotes with those from rs2312147 homozygotes. Additionally, to test any potential *cis*-effects associated with the expressed SNP rs1051061 itself, all corrected cDNA allele ratios were compared against all corrected genomic DNA allele ratios. All statistical comparisons were made by *t* tests using SPSS 20 for Mac (SPSS, Inc.). All *t* tests were

two tailed and where differences in variance were detected between comparison groups (Levene's test, $p<0.05$), a *t* test that assumes unequal variance was used.

4.2.9 Linkage disequilibrium and Haplotype analysis

Linkage disequilibrium (LD) between the schizophrenia risk SNP rs2312147 and the expressed SNP rs1051061 was assessed using Haploview 4.2 (Barrett et al., 2005). LD was calculated separately for the 96 post-mortem adult subjects and for the 40 first-trimester and 110 second-trimester foetal samples that were initially genotyped. Demographics of the individuals used for the LD analysis are shown in Table 4.6

Table 4.6 Demographics of individuals used to assess linkage disequilibrium between rs2312147 and rs1051061.

Brain sample	N	Sex		Age at Death			Post mortem delay (h)	
		M	F	Mean		Range	Mean	Range
Adult Brain	96	48	38	68.9	(17.7)	18-102	34.6 (18.1)	5 - 85
DLPFC				years		years		
Foetal Brain								
second trimester	110	54	51	15.4	(2.6)	13-26	-	-
	40	-	-	weeks		weeks	-	-
first trimester				10.1	(1.8)	5-12		
				weeks		weeks		

* N= the total number of subject used in the study. Sex: F=female, M= male. Numbers in parenthesis represent the standard deviation.

4.2.10 Bioinformatic analysis

All the genetic variants within the *VRK2* locus (chr. 2: 58.134.786 – 58.387.055, assembly hg19) were retrieved from 1000 genomes (Genomes Project et al., 2012) using tabix (Li, 2011). LD values between those variants and the expressed SNP rs1051061 and risk SNP rs2312147 were obtained using PLINK (http://pngu.mgh.harvard.edu/purcell/plink/) and the software SNAP (Johnson et al., 2008).

The variant effect predictor tool from ENSEMBL genome browser (http://www.ensembl.org/index.html) was used to extract the functional consequences of the variants in the *VRK2* region.

The software CONDEL (http://bg.upf.edu/condel/home) (Gonzalez-Perez and Lopez-Bigas, 2011) was used to evaluate the deleteriousness of non-synonymous variants in the *VRK2* region.

4.3 Results

4.3.1 Linkage disequilibrium

Linkage disequilibrium analyses were carried out between the expressed SNP rs1051061, used for the allelic expression analysis, and schizophrenia risk SNP rs2312147. LD analysis results for adult and foetal samples are shown in Figure 4.5 and 4.6.

The estimated frequency of the 4 haplotypes in the adult brain sample and the combined sample (adult brain and foetal brain sample) are shown in Table 4.7. Frequencies for the combined sample were obtained because Haploview did not generate the haplotype frequencies for the foetal sample alone (first and second trimester of gestation).

Haploview results indicated that the risk C-allele of rs2312147 was usually in phase with the G-allele of the expressed SNP rs1051061, and the non-risk T-allele of rs2312147 usually in phase with the A-allele of expressed SNP rs1051061, when a subject was heterozygous at both loci.

Table 4.7 Estimated frequency for all haplotypes for adult brain and combined sample.

	rs1051061	rs2312147	Frequency Adult brain sample	Frequency combined sample
Haplotype	G	C	0.420	0.349
	A	T	0.331	0.315
	A	C	0.234	0.315
	G	T	0.015	0.022

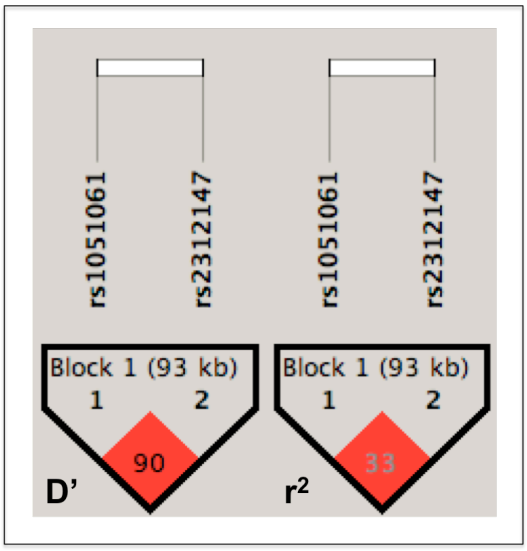


Figure 4.5 LD analysis of SZ risk SNP rs2312147 and the expressed SNP rs1051061 in Adult brain samples.

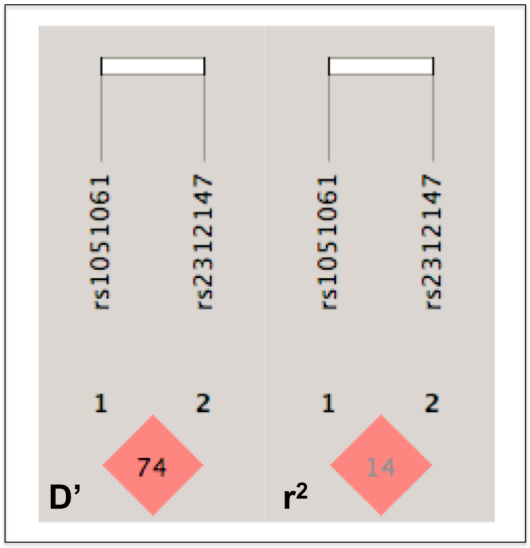


Figure 4.6 LD analysis of SZ risk SNP rs2312147 and the expressed SNP rs1051061 in Foetal brain samples.

4.3.2 Effect of schizophrenia risk SNP rs2312147 on *VRK2* allelic expression in adult brain

Allelic expression of VRK2 at the expressed SNP rs1051061 in the dorsolateral prefrontal cortex is illustrated in Figure 4.7. The effect of the schizophrenia associated polymorphism rs2312147 on the VRK2 allelic expression imbalance was tested by comparing homozygotes and heterozygotes for rs2312147. The average corrected cDNA allele ratio for individuals who were heterozygous for rs2312147 was 0.73 (range: 0.44–1.12) compared with an average ratio of 0.69 (range: 0.04 –1.11) for rs2312147 homozygotes. Allele ratios did not significantly differ between rs2312147 homozygous and heterozygous in adult DLPFC (P = 0.63). Conversely, significant differences were detected between cDNA and genomic DNA allele ratios in both rs2312147 homozygotes and heterozygotes ($P = 1.32*10^{-7}$, respectively). This suggests the presence of regulatory variants other than rs2312147 operating in this brain region, which are in strong LD with the expressed SNP rs1051061 and responsible for the allelic expression imbalance of *VRK2*.

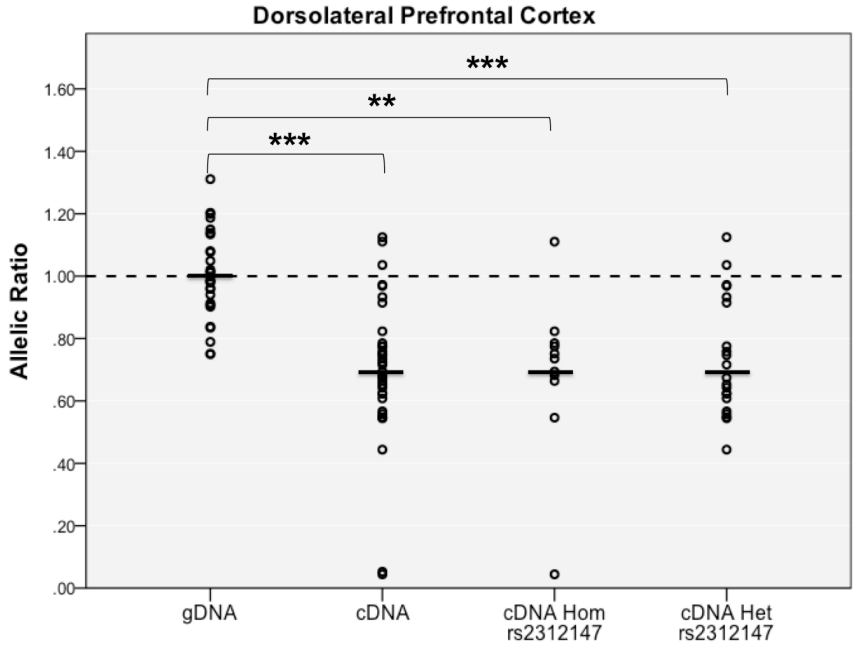


Figure 4.7 *VRK2* Allelic Expression in Adult Human DLPFC

Individual data points represent the average G/A allele ratio at the expressed single-nucleotide polymorphism (SNP) rs1051061 from four assays of genomic DNA (gDNA) or cDNA from each heterozygous subject. All gDNA ratios, cDNA ratios and cDNA ratios divided into rs2312147 heterozygotes and homozygotes are presented. The mean allele ratio is represented by a horizontal bar. The 1:1 ratio derived from the average cDNA ratio is indicated by a dotted line. There is no significant difference in cDNA allele ratios between homozygotes (N= 11) and heterozygotes (N=21) for the schizophrenia risk SNP rs2312147. cDNA allele ratios for heterozygotes and homozygotes for rs1051061 are significantly lower than gDNA allele ratios (P = $7.78*10^{-7}$ and P = 0.002, respectively)

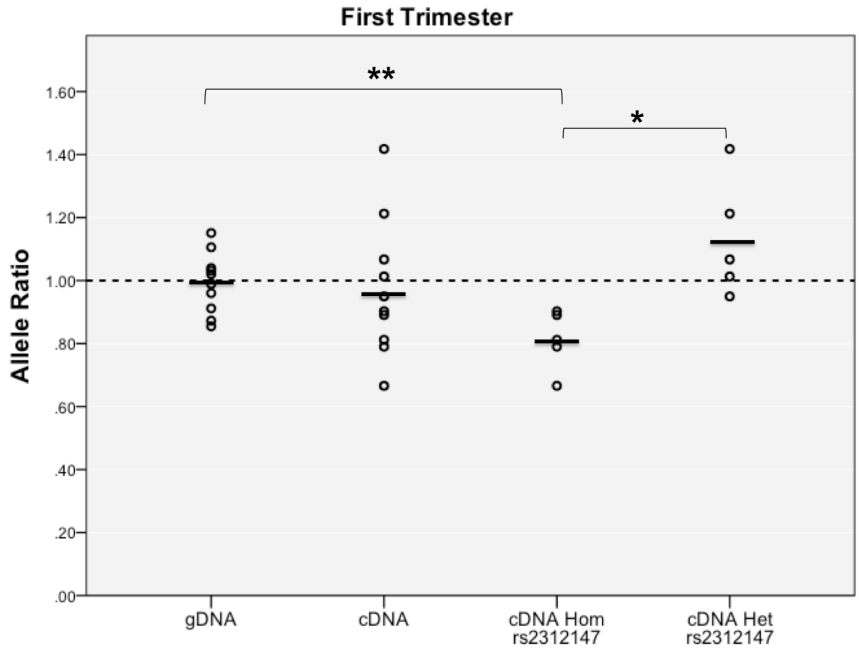
4.3.3 Effect of schizophrenia risk SNP rs2312147 on *VRK2* allelic expression in foetal brain

Since schizophrenia has been widely described as a neurodevelopmental disorder and due to the apparent lack of effect of rs2312147 in adult brain I decided to test the effect of rs2312147 on *VRK2* allelic expression in the human foetal brain. Whole brain samples from foetuses from first and second trimester of gestation were used to analyse allelic expression imbalance of VRK2 at the expressed SNP rs1051061 (Figure 4.8).

In first trimester foetal brain samples, allele ratios between rs2312147 heterozygous and homozygous did significantly differ (P = 0.015) with an average corrected cDNA allele ratio for individuals who were homozygous for rs2312147 of 0.81 (range: 0.67– 0.90) compared with an average ratio of 1.13 (range: 0.95 –1.42) for rs2312147 heterozygotes, suggesting a cis effect of the risk SNP from this time point of development.

Allele ratios in cDNA from rs2312147 homozygotes were significantly different from allele ratios in gDNA (P = 0.008). Allele ratios in cDNA from rs2312147 heterozygotes did not significantly differ from gDNA allele ratios (P = 0.180).

In common with first trimester brain tissue significant differences were observed according to rs2312147 genotype (P = 0.037) in second trimester brain tissue, averaging 0.92 (range: 0.65 – 1.25) in rs2312147 heterozygotes and 0.71 (range: 0.27– 0.91) in homozygotes. Allele ratios in cDNA from both rs2312147 homozygotes and heterozygotes were significantly different from allele ratios in gDNA (P = 0.009 and P = 0.045, respectively).



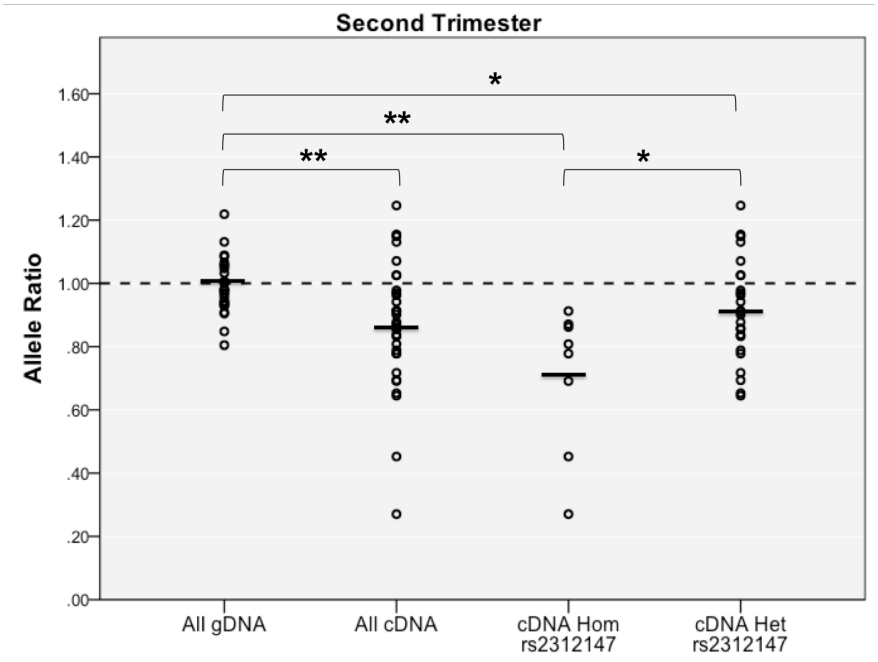


Figure 4.8 *VRK2* Allelic Expression in Foetal Human Brain

Individual data points represent the average G/A allele ratio at the expressed single-nucleotide polymorphism (SNP) rs1051061 from four assays of genomic DNA (gDNA) or cDNA from each heterozygous subject. All gDNA ratios, cDNA ratios and cDNA ratios divided into rs2312147 heterozygotes and homozygotes are presented. The mean allele ratio is represented by a horizontal bar. The 1:1 ratio derived from the average cDNA ratio is indicated by a dotted line. In the first trimester foetal brain, there is a significant difference in cDNA allele ratios between homozygotes (N= 5) and heterozygotes (N=5) for the schizophrenia risk SNP rs2312147. Allele ratios in cDNA from rs2312147 homozygotes were significantly different from allele ratios in gDNA (P = 0.008). Allele ratios in cDNA from rs2312147 heterozygotes did not significantly differ from gDNA allele ratios (P = 0.180) In the second trimester foetal brain, there are again significant differences in cDNA allele ratios between homozygotes (N= 8) and heterozygotes (N=25) for the schizophrenia risk SNP rs2312147 (P = 0.002). cDNA allele ratios for heterozygotes and homozygotes for rs1051061 are significantly lower than gDNA allele ratios (P = 0.045 and p= 0.009)

4.3.4 Bioinformatics analysis to explain allelic expression imbalance of *VRK2*.

The data described above indicate a *cis*-regulatory effect associated with genotype at schizophrenia SNP rs2312147 in foetal brain. However, as cDNA allele ratios in homozygotes were significantly different from gDNA allele ratios, it appears that there are additional *cis*-regulatory variants, active in all assayed tissues, operating on *VRK2*. As allele ratios in cDNA were generally reduced relative to those in gDNA, the regulatory variant is likely to be in strong LD with the expressed SNP (rs1051061) used to assay *VRK2* allelic expression. I therefore used the software SNAP (Johnson et al., 2008) and the variant effect predictor tool from ENSEMBL genome browser (http://www.ensembl.org/index.html) to identify SNPs at the *VRK2* locus that are in strong LD ($r^2 > 0.8$) with rs1051061 and have regulatory potential (Table 4.8).

Table 4.8 Variants in *VRK2* locus in high LD with expressed SNP rs1051061 and their predicted functional consequences.

SNP ID	r^2	D'	Consequence
rs1051061	1	1	missense_variant, 3_prime_UTR_variant,NMD_transcript_variant,downstream_gene_variant
rs1016771	1	1	intron_variant,NMD_transcript_variant, regulatory_region_variant
rs13011711	1	1	intron_variant,NMD_transcript_variant, regulatory_region_variant
rs3771213	0.961	1	intron_variant,NMD_transcript_variant
rs12999901	0.922	1	intron_variant,NMD_transcript_variant
rs13033536	0.922	1	intron_variant,nc_transcript_variant
rs35248619	0.922	0.96	intron_variant,NMD_transcript_variant
rs57390191	0.922	0.96	intron_variant,NMD_transcript_variant
rs62140014	0.922	0.96	intron_variant,NMD_transcript_variant
rs11678320	0.883	0.958	intron_variant,NMD_transcript_variant

Additional bioinformatic analysis indicated that the expressed SNP rs1051061 encodes an amino acid change (Isoleucine to valine, aminoacid 167) in the VRK2 protein, within the Serine/Threonine kinase domain. Condel predicts that this change is likely to have a functional effect on protein function. The analyses showed that the expressed SNP rs1051061 is a missense variant. I used the software Condel to find out it deleteriousness score, the results are in the table below:

Table 4.9 Condel scores for rs1051061.

Location	Allele	Gene	Protein position	Aminoacid change	Sift	PPH2	Condel score	Condel label
2:58316814	G	VRK2	167	I/V	0.04	0.947	0.74	deleterious

Condel score ranges from 0 to 1, with 0.0 = neutral and 1.0 = deleterious; SIFT: Ranges from 0 to 1, being <= 0.05 predicted to be damaging and > 0.05 tolerated; PPH2: Ranges from 0 to 2, being <= 0.05 neutral and > 0.05 probably damaging.

4.4 Discussion

The main aim of the experiments described in this Chapter was to assess allele-specific expression of *VRK2* and determine if a SNP associated with schizophrenia (rs2312147, (Steinberg et al., 2011)) is also associated with altered *cis*-regulation of *VRK2*. *VRK2* allelic expression was assessed in adult DLPFC and whole brain from two stages of foetal development.

The general distortion in cDNA allele ratios at expressed SNP rs1051061 suggests the presence of *cis*-acting influences affecting *VRK2* expression in the adult and foetal brain. However, as homozygotes for risk SNP rs2312147 show distortion in cDNA allele ratios, it would appear that rs2312147 cannot account for this apparent *cis*-effect. An interesting possible explanation, albeit speculative without further evidence and assuming that the findings are not due to differences in reverse transcription efficacy between the two alleles, is that the observed allelic imbalance is due to either the expressed SNP itself or to one in high LD with it. The expressed SNP is non-synonymous and according to ENCODE data localizes within a nonsense mediated decay (NMD) site. It is therefore possible that G-allele appears as less expressed due to NMD. Additionally, the expressed SNP is in high LD with several SNPs that are in potentially regulatory regions, like rs13011711 (r^2 = 1) localized in the middle of DNAse hypersensitivity site and transcription binding site, according to ENCODE data (Table 4.8). Further analyses will be necessary to rule out or establish the role of these variants.

The effect of rs2312147 on *VRK2* allelic expression was initially tested in the DLPFC of the adult brain. This region was chosen because a large number of RNA samples from this brain region were already available within Dr Bray's lab and it has been strongly implicated in schizophrenia pathophysiology (Berman et al., 1988). There was no significant effect of genotype at rs2312147 on *VRK2* allelic expression in adult DLPFC. However, it is also important to consider that effects of *cis*-regulatory variation can differ across brain regions (Buonocore et al., 2010), and potentially across cell types. Hence,

effects of rs2312147 on *VRK2* allelic expression in other adult brain regions cannot be excluded.

Genotype at rs2312147 had a significant effect on *VRK2* allelic expression in foetal brain samples from both the 1st and 2nd trimester of gestation. However, in both 1st and 2nd trimester tissue, heterozygosity for rs2312147 was associated with *less* allelic expression imbalance of *VRK2*, with the risk (C) allele increasing the expression of the allele of the expressed SNP that is otherwise under-expressed. A possible explanation for how this could influence schizophrenia risk would be that the *cis*-effects of rs2312147 alter the functional impact of the deleterious amino acid encoded by the expressed SNP rs1051061. Lappalainen et al. (Lappalainen et al., 2011) present data to suggest strong selection pressure against deleterious amino acids occurring on high expression haplotypes. They further report an enrichment of putative epistatic effects between deleterious coding variants and eQTL associated with common diseases by GWAS.

The association between genotype at a schizophrenia risk variant and *VRK2* allelic expression in foetal brain suggests a molecular risk mechanism for schizophrenia operating during early brain development. Schizophrenia is widely hypothesized to be a neurodevelopmental disorder (Harrison, 1997; Weinberger, 1995), but direct molecular evidence for this is lacking. Genome-wide significant risk variation for schizophrenia has recently been reported to affect *ZNF804A* expression specifically in the foetal brain (Hill and Bray, 2012). However, whereas that study found that the effects on *ZNF804A* expression were specific to 2nd trimester tissue, the effects on *VRK2* appear to start even earlier, during the 1st trimester of foetal development. This study also provides evidence for genetic mediated risk for schizophrenia operating long before the evident appearance of the disease.

Chapter 5 Effects of *VRK2* manipulation in human neural progenitor cells

5.1 Introduction

The *VRK2* gene is located at chromosome 2p16.1, encodes a member of the vaccine-related kinase family (VRK) serine/threonine kinases. Alternative splicing generates two known isoforms, A and B, of 508 and 397 amino acids in length, respectively. VRK2A has been shown to localise to the cytoplasmic side of the endoplasmic reticulum and mitochondrial membrane by virtue of a hydrophobic tail in the C-terminal region (Blanco et al., 2006). VRK2B is shorter and is localised to the cytosol as well as the nucleus and outside the nucleolus (Blanco et al., 2006). Subcellular localization of VRK2 A and B is dependant on cell type (Blanco et al., 2006).

Genetic variation at the *VRK2* locus has been associated with schizophrenia and the SNP rs2312147 was identified as the strongest statistical signal (Irish Schizophrenia Genomics and the Wellcome Trust Case Control, 2012; Steinberg et al., 2011). However, the mechanism behind this association is unknown. To date there is only one study from Li et al., (Li et al., 2012) reporting an up-regulation of VRK2 mRNA expression in schizophrenic patients, but they did not find a significant association between the risk SNP rs2312147 and VRK2 mRNA expression.

VRK2 has been associated to epilepsy (Consortium et al., 2012) and a microdeletion syndrome involving chromosome 2p15-p16.1 (Chabchoub et al., 2008; de Leeuw et al., 2008; Felix et al., 2010; Liang et al., 2009; Rajcan-Separovic et al., 2007). To date, most reported functional studies on VRK2 are concerned with its involvement in cancer, as it has been linked to breast cancer via its regulation of ErbB2-MAPK signalling (Blanco et al., 2006; Fernandez et al., 2010). Given the evidence for an effect of schizophrenia risk variant rs2312147 on VRK2 expression in foetal brain reported in chapter 4, it is

important to characterise the functional role(s) of VRK2 in the developing human brain. However, there are no published data on this to date.

The VRK family of serine/threonine kinases are upstream of several transcription factors and form part of multiple signalling pathways. In non-neural tissues they are known to be involved in several cellular processes such as proliferation, apoptosis and response to stress (Blanco et al., 2006; Blanco et al., 2007; Monsalve et al., 2013). *VRK2* expression has been detected in many tissues (Lukk et al., 2010; Nezu et al., 1997), and several different human brain regions (Allen Brain Atlas, (Hawrylycz et al., 2012)). RNA sequencing projects suggest the long but not the short isoform may be present in brain tissue (Illumina Human Bodymap 2.0 Data displayed at www.ensembl.org). This dataset represents a first approximation and additional surveys will help clarify the relevance of different isoforms for brain function. In brain tissue, VRK2 has a marked developmental expression pattern with high expression in the first weeks of embryonic development and then decreasing expression (no data available from 24 to 40 weeks). VRK2 later shows increasing expression during childhood and a more stable pattern from 10 years onwards (data available up to 80 years) (Figure 5.1). A more detailed inspection suggests that this pattern is consistent for different brain regions with the exception of the cerebellum, where it does not show much variation in expression throughout life (Figure 5.2).

Interestingly, Tebbenkamp et al., (Tebbenkamp et al., 2014) summarised the developmental brain expression pattern of the human brain and found that genes associated with broad biological functions follow distinct expression patterns. During embryonic development VRK2 follows a pattern associated with genes involved in cell proliferation and after birth of those associated with myelination, dendrite development and (less so), synaptic development. It does not; however, seem to adhere to the expression pattern of genes associated with progenitors and immature neurons (Figure 5.3). VRK2 shows high expression under conditions of cell proliferation, such as early brain development (Figure 5.2), and in highly proliferative tissues such as testis, thymus and foetal brain (Nezu et al., 1997).

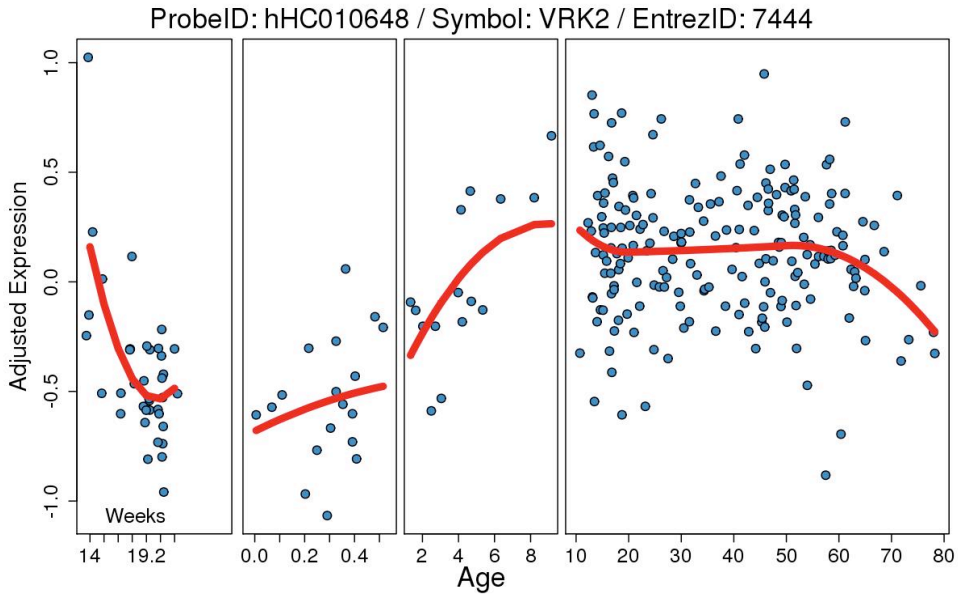


Figure 5.1 Expression of *VRK2* in prefrontal cortex samples ranging from 14 post-conception weeks to 80 years of age. Data was generated by the National Institute of Mental Health / Lieber Institute for Brain Development, using spotted microarray and accessed via their publicly available database at http://braincloud.jhmi.edu, (Colantuoni et al., 2011).

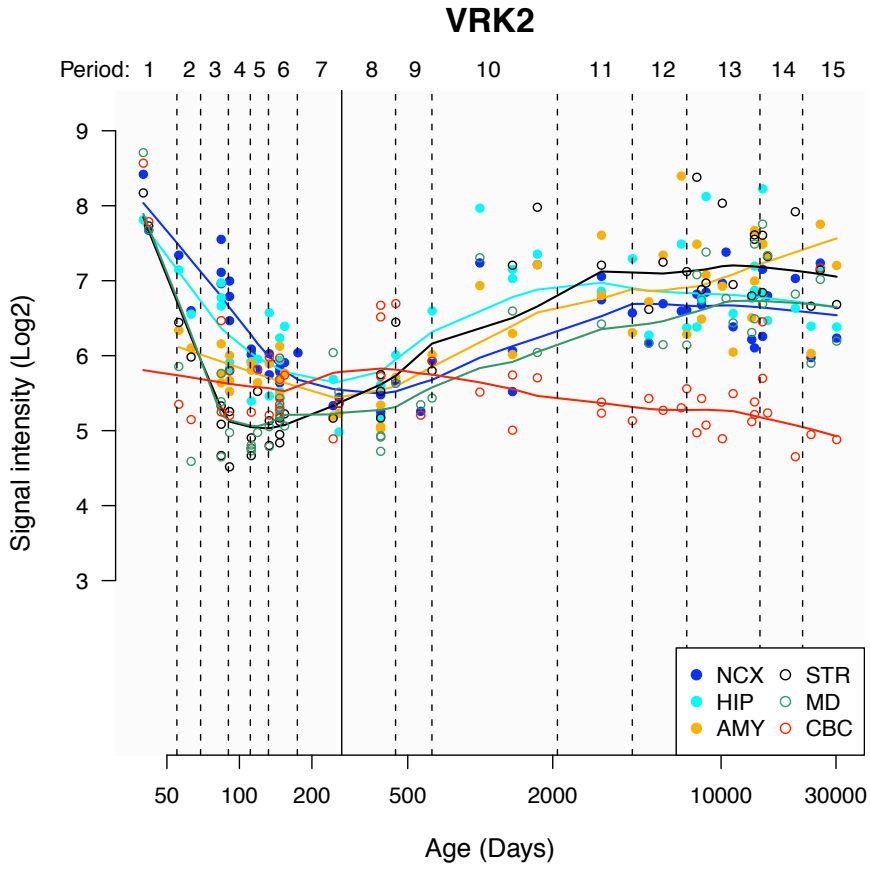


Figure 5.2 Expression of VRK2 in 6 human brain regions ranging from 6 post-conception weeks to 82 years of age. Data were generated by the Department of Neurobiology, Yale University School of Medicine, using Affymetrix GeneChip Human Exon 1.0 ST Arrays and accessed via their publicly available database at *http://hbatlas.org*. NCX= neocortex; HIP= hippocampus; AMY= amygdala; STR= atriatum; MD= mediodorsal nucleous of the thalamus; CBC= cerebellar cortex (Johnson et al., 2009; Kang et al., 2011).

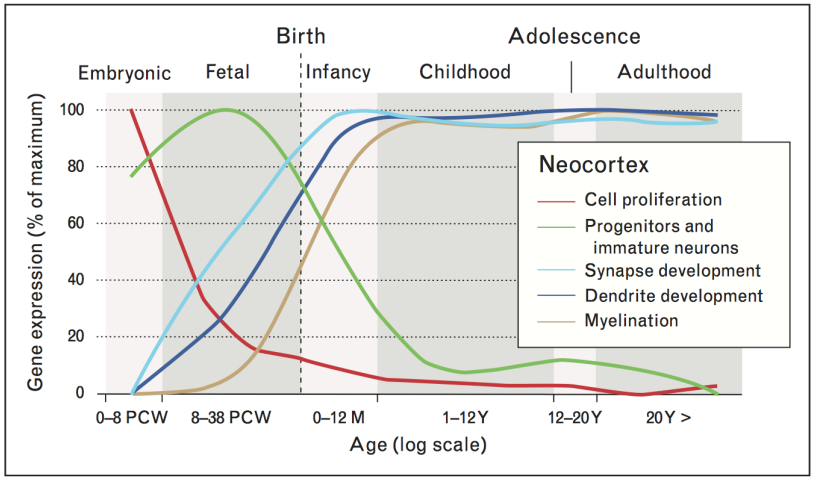


Figure 5.3 Expression trajectories of genes involved in neurodevelopmental processes ranging from early brain development until adulthood (taken from Tebbenkamp et al., (Tebbenkamp et al., 2014))

VRK2A and B isoforms are able to undergo autophosphorylation on their Thr residues (Nichols and Traktman, 2004) and both of them are able to phosphorylate p53 at Thr-18 modulating its stability (Blanco et al., 2006), and c-Jun (Sevilla et al., 2004). Other known substrates are the transcription factor nuclear factor of activated T-cell (NFAT1), involved in cell invasion (Vazquez-Cedeira and Lazo, 2012), the barrier to autointegration factor (BAF), which plays an important role as a regulator of nuclear architecture during the phases of the cell cycle (Nichols et al., 2006) and coilin (COIL), a nuclear protein involved in neurological conditions (Sanz-Garcia et al., 2011).

Protein-protein interaction studies have shown that VRK2 proteins are regulated by their interaction with Ran GTPase (Sanz-Garcia et al., 2008). Additionally, these studies have shown that the VRK2A isoform; anchored to the endoplasmic reticulum, interacts directly with scaffold proteins that regulate cellular signalling pathways. It has been shown to interact with JIP1, a scaffold protein that assembles mitogen-associated protein (MAP) kinase signalling complexes (Blanco et al., 2008), and this interaction can modulate the magnitude of the stress response to hypoxia and to interleukin-β (Blanco et al., 2007; Blanco et al., 2008). VRK2 also interacts with the kinase suppressor of RAS1 (KSR1) scaffold protein, reducing the signal mediated by MAPK in response to ErbB2 in breast cancer. Additionally, VRK2 interacts with MEK, a kinase suppressor of Ras 1, resulting in ERK signalling pathway inhibition (Fernandez et al., 2010; Fernandez et al., 2012). VRK2 also appears to regulate apoptosis by its interaction with BHRF (Epstein-Barr virus) and with Bcl-xL (Li et al., 2006; Monsalve et al., 2013). Furthermore, VRK2 interacts with the eukaryotic chaperonin TCP-1 ring complex (TRiC) leading to its ubiquitination through activation of COP1 E3 ligase activity. In turn this has been shown to lead to poly-Q-aggregation and cytotoxicity involved in neurodegenerative disorders such as Huntington's disease (Kim et al., 2014).

Overall, based on its expression pattern, functional studies and the literature associated with its interactions, VRK2 is likely involved in the regulation of multiple cellular signalling pathways controlling cell proliferation, apoptosis and response to stress. Its expression in

brain development (especially early on) and its involvement in pathways linked to neurological disorders support an important function in the CNS. However, no studies of VRK2 function in brain have been reported to date.

5.1.1 The aim of the present study

The aim of the experiments presented in this chapter was to explore the molecular and cellular consequences of VRK2 manipulation in cells derived from human foetal brain.

To this end I used a combination of siRNA, genome-wide expression profiling, pathway analysis and automated assays of cell proliferation.

5.2 Methodology

5.2.1 Human neuronal progenitor cell line

Experiments were carried out in the conditionally immortalised neuronal progenitor cell line CTX0E16. The cell line CTX0E16 was obtained from ReNeuron Ltd (www.reneuron.com) under a Material Transfer Agreement. It was derived from first trimester human foetal cortical neuroepithelium (Pollock et al., 2006). The cells have been conditionally immortalised by genomic incorporation of the *c-myc*ER^{TAM} transgene, which can be conditionally activated by 4-hydroxy-tamoxifen (4-OHT) to stimulate proliferation (Littlewood et al., 1995).

A detailed description of cell culture protocol followed in the present study has been described in Chapter 3 section 3.2.1.

5.2.2 Transfection of siRNA and DNA vectors with a nucleofection system

Transfection of siRNAs and DNA vectors was performed by Nucleofection^{TM} technology from Amaxa® (now Lonza, Gaithersburg, MD), using the Cell Line Nucleofector® Kit V (VACA-1003) and the Nucleofector™ Device from Amaxa® (now Lonza, Gaithersburg, MD). Nucleofection^{TM} technology is an electroporation method developed by Lonza to efficiently delivery siRNAs and plasmids into cells (Gresch et al., 2004).

RNAi was achieved using pre-designed Stealth RNAi™ (Invitrogen), which uses next-generation RNAi chemistry to prevent off-target effects. Two non-overlapping Stealth RNAi™ siRNA for *VRK2* (siRNA1: HSS111310 and siRNA2: HSS111311) were transfected into separate flasks of seeded cells alongside a Stealth RNAi^{TM} siRNA negative universal control (Invitrogen), and a BLOCK-iT™ Alexa Fluor® Red Fluorescent Oligo to assess the transfection efficiency.

The sequences of the siRNA targeting *VRK2* are shown in Table 5.1. Both siRNAs target sequence present in exon 5 and therefore target all *VRK2* transcripts.

Table 5.1 Details of siRNAs used for *VRK2* knockdown.

siRNA ID	Oligo	Sequence	*VRK2* exon
siRNA1	HSS111310	GGAUAGAACGCAAACAACUUGAUUA	5
siRNA2	HSS111311	UGGAUCUGGUCUGACUGAAUUCAAG	5

Overexpression of *VRK2* was attempted using the mammal expression vector pCMV-SPORT6-VRK2 (pVRK2) (Thermo Scientific, MHS6278-202801669). To monitor transfection efficiency the pmaxGFP® Vector (Lonza) was transfected alongside pVRK2. A map for each vectors is shown in Figures 5.4 and 5.5.

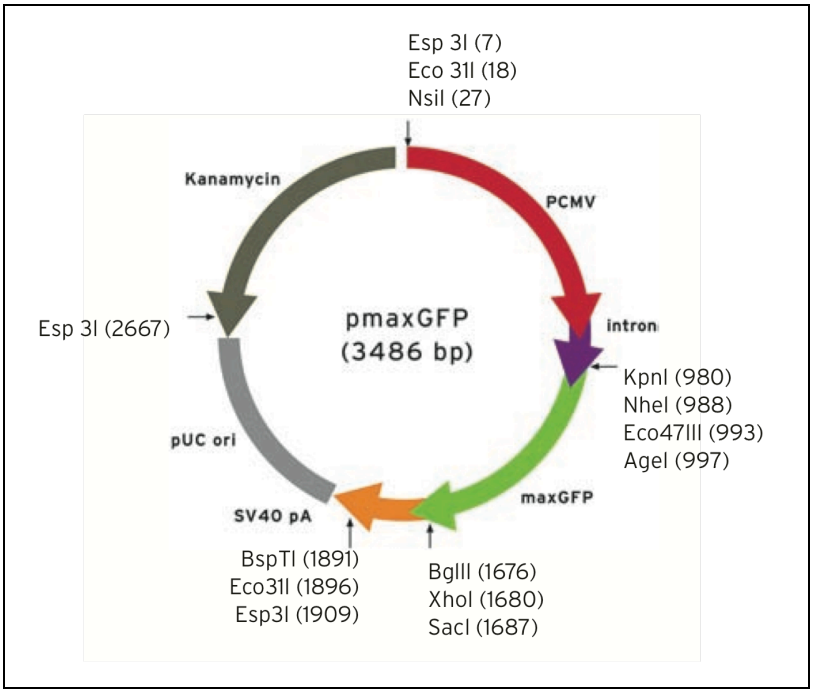


Figure 5.4 Map of pmaxGFP vector.

Positive control vector pmaxGFP® encodes the green fluorescent protein (GFP) from copepod Pontellina p. Cells expressing maxGFP® can be examined by fluorescence microscopy to check transfection efficiency. (Image taken from www.Lonza.com)

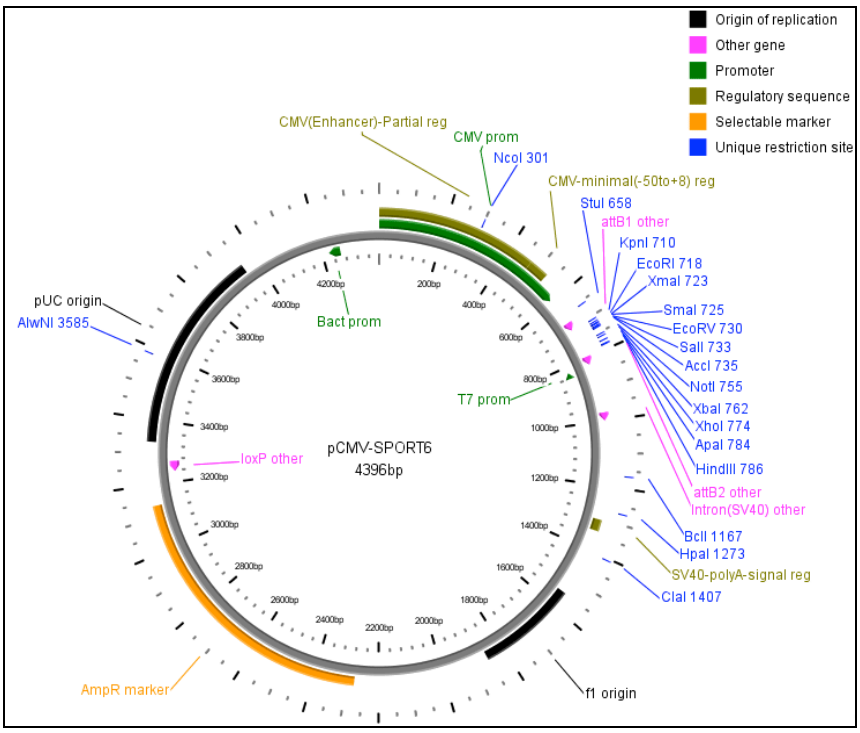


Figure 5.5 Map of overexpression vector for VRK2, pCMV-SPORT6.

The pCMV-SPORT-VRK2 construct contained the full cDNA sequence of *VRK2* clone BC027854. *VRK2* sequence was inserted between *Not*I and *Eco*RV restriction sites (Image taken from www.thermoscientific.com).

CTX0E16 cells used for transfection were expanded in laminin-coated T75 flask to 80% confluence, as the nucleofection protocol requires cells in their logarithmic growth phase. 24 hours before transfection the RMM media was replaced with 10 mL of RMM media without 4-OHT (represented as RMM ++-). 4-OHT was excluded so that proliferation was not artificially stimulated by *c-myc* overexpression. CTX0E16 cells are able to retain their nestin-positive, neuronal progenitor state in the presence of growth factors bFGF and EGF (Dr G. Anderson and Dr B. Williams, personal communication). Separate flasks of CTX0E16 cells were transfected with either Stealth siRNA1, Stealth siRNA2, Stealth siRNA™ negative universal control, positive control BLOCK-iT™ Alexa Fluor® Red Fluorescent oligo, pVRK2 vector (Thermo Scientific) or the pmaxGFP® Vector (Lonza).

A detailed Nucleofection procedure is described below:

Cells were harvested by addition of Accutase (Invitrogen) and centrifugation at 900 rpm for 5 minutes at room temperature. The pelleted cells were washed once with DPBS, to eliminate any trace of media that could interfere with transfection, and centrifuged again at 900 rpm for 5 minutes. The final cell pellet was resuspended in 100 mL of nucleofection® sample, containing siRNA or DNA vector and transferred to an Amaxa™ certified 100 µL aluminum electrode cuvette. Details of reagents used per each transfected flask are shown in Table 5.2.

Each sample was processed separately to avoid storing the cells for more than 15 minutes in Nucleofector® solution.

The cuvette was inserted into the Nucleofector™ Device and the program T-030 was selected for nucleofection. The cuvette was then rinsed with 500 µL of pre-warmed RMM (++-) media (using the provided pipette) and the suspension was transferred to a laminin-coated flask with 9.5 mL of warmed media and incubated in humidified 37 °C/5% CO_2 incubator until analysis.

Table 5.2 Contents of one Nucleofection® Sample.

Reagent	Concentration	Final volume added
CTX0E16 cells	$5*10^6$ cells	-
1 - 5 µg pVRK2 (in max. 5 µL)	2 µg/ µL	2 µL
or 2 µg pmaxGFP®	0.5 µg/ µL	4 µL
or 2 nM-2 µM siRNA	20 µM	5 µL
Nucleofector® solution containing: **18 µL stabilizer supplement and** **82 µL Nucleofector™ Solution**		100 µL

For proliferation assays, 50 µL of suspended cells was taken from the transfected flask and transferred to a well in a Nunc-Immuno™ MicroWell™ 96 well solid plate. This was repeated for each treatment with a total of 10 wells per treatment.

In order to estimate transfection efficiency, cells transfected with the BLOCK-iTTM Alexa Fluor® Red Fluorescent oligo and the pmaxGFP® Vector were checked 24 hours post-transfection by fluorescence microscopy.

Media was replaced at 48 hours post-transfection with RMM (++-). At 120 hours post-transfection cells from each T75 were separately harvested by addition of Accutase. Resuspended cells from each treatment were divided into 2 Falcon tubes and centrifuged at 900 rpm for 5 minutes. One of the cell pellets was used for RNA extraction to which 500 µL of Tri-Reagent® Solution (Ambion, Life Technologies, Grand Island, N.Y.) was added. The second pellet was used for protein extraction. Both tubes were stored at -80°C. An overview of the transfection time-line is represented in Table 5.3.

Cells in proliferation plates were fixed with 4% Paraformaldehyde (PFA). A detailed protocol for the proliferation assays is described in detail below this section.

The whole protocol for the six treatments (siRNA1, siRNA2, Stealth siRNA^{TM} negative universal control, BLOCK-iT^{TM} Alexa Fluor® Red Fluorescent oligo, pVRK2 and the pmaxGFP® Vector) was repeated four times to give four biological replicates for each condition.

Table 5.3 Time-line protocol for *VRK2* knockdown and overexpression assays.

Day 0	**Replace RMM media for 10 mL of RMM ++-**
Day 1	Transfect: • 1 flask with siRNA 1 • 1 flask with siRNA 2 • 1 flask with negative control • 1 flask with red oligo • 1 flask with pVRK2 • 1 flask with pmaxGFP® Transfer 50 uL suspension cell from each flask (each treatment) to 10 wells of a 96 well plate.
Day 2	Check florescent cell in flask transfected with red-oligo and pmaxGFP®
Day 3	Replace media to all flasks and 96 well plate with RMM ++- (10 mL per flask and 50 µl per well)
Day 4	-
Day 5	Collect cells from the different treatments for RNA extraction and protein extraction.

5.2.3 RNA extraction and cDNA synthesis

Pelleted cells were separately homogenized in 500 µL of Tri-Reagent® Solution (Ambion, Life Technologies, Grand Island, N.Y.). Total RNA was extracted following the protocol described in section 3.2.3 (Chapter 3). RNA concentration and purity was quantified using the NanoDrop™ 1000 spectrophotometer. RNA integrity was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) following manufacturer protocol. All samples had an RNA integrity number (RIN) >9.4.

DNAse treatment of total RNA samples was performed with TURBO DNA-*free*™ (Ambion) (as described in section 3.2.4, Chapter 3). cDNA was synthesized from 1 µg total RNA from each sample using SuperScript™ III reverse transcriptase (Invitrogen) and

random decamers (Ambion) in final reaction volume of 20 µL (a detailed protocol can be found in section 3.2.6, Chapter 3). The resulting cDNA samples were diluted to the "standard cDNA concentration" (1/7), by adding 120 µL of nuclease-free water (Sigma-Aldrich) to each 20 µL of cDNA.

5.2.4 Quantitative PCR (qPCR)

Quantitative PCR was used to assess VRK2 knockdown and overexpression. qPCRs were completed on 96-well plates using an MJ Chromo4™ Real Time PCR Detector (Bio-Rad) and MJ Opticon Monitor analytic software (Bio-Rad). Reactions were carried out in a total volume of 20 µL, containing diluted cDNA, primers at 200 nM and 1x HOT FIREPol® EvaGreen® qPCR Mix Plus (Solis BioDyne). Table 5.4 shows the volume and concentration of the reagents used. The cycling programme used for qPCR reactions can be found in Table 5.5.

Table 5.4 Details of reagents used for each qPCR reaction.

Reagent	Supplier	Concentration	Volume (µL)
HOT FIREPol ® EvaGreen ® qPCR Mix Plus	Solis BioDyne	5x	4
Forward and reverse primer mix	Sigma-Aldrich	2 µM	2
Nuclease-free water		-	6
cDNA template/nuclease-free water	-	-	8
Total	-	-	20

Table 5.5 Typical qPCR cycling conditions.

Step	Temperature (°C)	Time (minutes)	Number of Cycles
Initial denaturation	95	15	1
Denaturation	95	0.3	44
Annealing	60	0.30	
Extension	72	0.3	
For the melting profile, the temperature was raised from 60 to 95 and the fluorescence **read every 1°C** increase in temperature with a hold time of 10 sec.			

QPCR primers were design using the Primer3Plus program available online (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/). Sequences of the primers used for qPCR are shown in Table 5.6.

Table 5.6 Detail of primers sequences used for qPCR analysis.

Gene	Primer	Sequence (5' - 3')
VRK2	Forward	AGCAAGTCAACAAGGCACAC
	Reverse	CCAATCAGTTTCTCCTCTTTC
GAPDH	Forward	CCTGACCTGCCGTCTAGAAA
	Reverse	ATCCTGGTGCTCAGTGTAGCC
RPL13A	Forward	CCTGGAGGAGAAGAGGAAAGAGA
	Reverse	TTGAGGACCTCTGTGTATTTGTC
CFDP1	Forward	CTTCACCAATCCCCTCTTCC
	Reverse	TGCCAAGAAGCAGAAAATGA

Four technical replicate qPCR reactions were performed to measure *VRK2* and the internal control genes *RPLI3A*, *GADPH* and *CFDP1* for each cDNA sample. QPCRs for *VRK2* and the internal controls were carried out on the same PCR plate for every transfection experiment and each treatment (siRNA1, siRNA2, Stealth siRNA™ negative universal control, pVRK2 and the pmaxGFP® Vector). The level of expression of each gene was measured against a relative standard curve created by serial dilution of pooled cDNA taken from each of the assayed samples. The following arbitrary values were given to standard curve: "5000" (undiluted), "1000", "200", and "40", according to their relative dilutions. Consequently, a relative value was obtained for each of the four replicate reactions for each cDNA sample.

5.2.5 Statistical analysis for qPCR

Description of the data:

1. Transfection experiments were done in four biological replicates. Each biological replicate has five treatments, siRNA1, siRNA2, Stealth siRNA™ negative universal control, pVRK2 and the pmaxGFP® Vector, leading to four biological replicates per treatment.
2. cDNA was obtained from RNA for the four biological replicates for each treatment. This gave 20 cDNA samples in total.
3. qPCR was used to measure the expression of *VRK2* and three internal control genes, *RPLI3A*, *GADPH* and *CFDP1*. This was performed four times, leading to four technical replicates for each of the 20 cDNA samples. qPCR for *VRK2* and the internal controls was carried out on the same PCR plate.
4. For each technical replicate I obtained one normalized value for *VRK2* per treatment per biological replicate. Normalization was calculated using the geometric normalization method (Vandesompele et al., 2002). Thus, from each technical replicate I obtained one normalized value for *VRK2* for each treatment per biological replicate. In total I obtained four technical replicates for four biological replicates each including the five treatments. These normalized values were used for statistical analyses.

The objective of the statistical analyses was to identify expression changes on *VRK2* mRNA level between control siRNA and *VRK2* siRNA conditions and pmaxGFP and pVRK2 conditions. To this end I used a linear mixed model including treatments as a fixed effect and biological replicates as random effects. I used a linear mixed model because it considers the systematic variation between biological replicates. Statistical analysis was performed using a linear mixed model in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.) Differences of $P < 0.05$ were considered significant. The exact command using R was:

lmer(log(normalized expression) ~ 1 + treatment + (1|biological_replicate),data=qpcr.m).

5.2.6 Total protein extraction and quantification for VRK2 knockdown.

Total protein was extracted from cells transfected with siRNA1, siRNA2 and Stealth siRNA™ negative universal control. Cells were lysed with the Pierce IP Lysis Buffer (25mM Tris-HCl pH 7.4, 150mM NaCl, 1mM EDTA, 1% NP-40 and 5% glycerol) (Thermo Scientific) containing a protein inhibitor cocktail (Thermo Scientific). 300 µL of lysing buffer was added to each sample separately and the mixture was transferred to a pre-chilled microcentrifuge tube and agitated in a rotary mixer for 30 minutes at 4 °C. The lysates were then clarified by centrifugation at 10.000 rpm for 10 minutes at 4 °C. The supernatant was then transferred to a clean pre-chilled microcentrifuge tube and store at -80 °C.

Quantification of total protein was achieved by using the Novagen® BCA Protein Assay Kit (Merck Millipore, 71285), following the micro-scale assay according to the manufacturer's instructions.

5.2.7 SDS-PAGE and inmunobloting

5 µg of total protein from each sample was separated on a 8% SDS/polyacrylamide gel and transferred to a polyvinylidene difluoride (PVDF) membrane in western transfer buffer (10% methanol, 25mM Tris-HCl, 190 mM glycine, pH 8.5) at 100 Volts for 1 hour at 4 °C. The membrane was then incubated in blocking buffer (3% (w/v) BSA in Tris-buffered saline with Tween-20 (TBST buffer: 50 mM Tris, 150 mM NaCl, 0.05% Tween-20, pH 7.5)) for 1 hour at room temperature. The membrane was then incubated overnight at 4°C with the primary antibody anti-VRK2 (Sigma-Aldrich, WH0007444M1) diluted 1:500 in blocking buffer. The blot was then washed three times with TBST buffer for 20 minutes and subsequently incubated for 1 hour at room temperature with the horseradish peroxidase-conjugated secondary antibody, goat anti-mouse (Pierce) diluted 1:20000. Finally, the membrane was washed three times with TBST buffer for 20 minutes and proteins were detected by using the SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific), according to the manufacturer's instructions.

Afterwards, the membrane was stripped with the Fluorescent WB striping buffer (Thermo Scientific) and incubated with primary antibodies for housekeeper proteins applied in blocking buffer at the following dilution: anti-GAPDH (AbD Serotec) 1:2000, anti-β-actin (Thermo Scientific) 1:2000 and anti-α−Tubulin (Genetex) 1:5000, for 1 hour at room temperature. The membrane was then washed three times with TBST buffer for 20 minutes and then incubated for 1 hour at room temperature with the horseradish peroxidase-conjugated secondary antibody, goat anti-mouse (Pierce) diluted 1:10000. Finally, the membrane was washed three times with TBST buffer for 20 minutes and proteins were detected by using the SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Scientific), according to the manufacturer's instructions.

Protein levels were quantified by measuring density of the band in ImageJ (Schneider et al., 2012).

5.2.8 Proliferation assay

Proliferation assay was performed following the protocol described in section 3.2.3 (Chapter 3).

5.2.9 Statistical analysis for cell proliferation

Data analysis

1. Transfection experiments were done in four biological replicates. Each biological replicate has three treatments, siRNA1, siRNA2 and Stealth siRNA™ negative universal control, leading to four biological replicates per treatment. For proliferation assays 10 wells of a Nunc-Immuno™ MicroWell™ 96 well solid plate, containing CTX0E16 cells, were transfected for each siRNA treatment, leading to 10 technical replicates for each biological replicates for each cell line.
2. The effect of *VRK2* knockdown on the proliferation of CTX0E16 cells was analysed by ki67 inmunostaining. Cell counts for proliferating cells (ki67 positives) and total number of cells per well for each treatment in the four biological replicates (four plates) were obtained from Cellinsight. These values were used in the statistical analyses.

The objective of the statistical analyses was to identify the differences in cell proliferation between cells treated with control siRNA and *VRK2* siRNA conditions. To this end I used a Poisson linear mixed model in R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.). This model compares the rate between treatments, taking into account the sample size, including cell counts for proliferating cell and total number of cells per treatment (siRNA1, siRNA2 and negative control siRNA) in the four biological replicates. The analysis considers the systematic variation between plates with a fixed effect per treatment and random effect per plate. Differences of $P < 0.05$ were considered significant.

The exact command using R was:

glmer(count_proliferating_cells ~ offset(log(count_total_cells)) + Treatment + (1|Plate) ,data=data.m2 , family='poisson').

5.2.10 Genome-wide gene expression profiling

After confirming significant *VRK2* knockdown, by qPCR, in both *VRK2* siRNA conditions (siRNA1 and siRNA2) compared against the control condition (Stealth siRNA™ negative universal control), the 12 RNA samples obtained from cells transfected with siRNAs (4 for siRNA1, 4 for siRNA2 and 4 for the negative control) were sent to the SGDP MRC centre, Institute of Psychiatry, King's College London, for microarray analysis. Genome-wide gene expression profiling was performed using the HumanHT-12 v4 Expression BeadChip (Illumina).

Microarray data was analysed using Bioconductor in R. First the lumidat Bioconductor package (M.J. Cowley, M. Pinese, D. Eby, M.R. Reich, A.V. Biankin, J. Wu, 2011. IlluminaIDatReader: A program for reading raw Illumina Gene Expression data (manuscript in prepration)) (https://github.com/drmjc/lumidat) was used to pre-process Illumina gene expression idat files, the format in which the data was received from the

SGDP centre. Then the lumi Bioconductor package (Du et al., 2008) was used to perform quality control, variance stabilization, normalization and gene annotation analysis.

5.2.11 Identification of differentially expressed genes and bioinformatic analyses.

Genes differentially expressed as a result of *VRK2* knockdown were identified using the limma Bioconductor package (Smyth, 2004).

To limit false results arising from low expression genes, probes that were not detected in at least all four samples of one of the comparison conditions with a detection *P* value > 0.95 were excluded. Microarray probes showing differences in expression between each *VRK2* siRNA and the negative siRNA condition were identified based on moderated t-test using limma on normalized microarray data. To identify genes differentially expressed I set a multiple-testing correction threshold of FDR < 0.05 and to select genes for additional analyses a nominally significant threshold of $P < 0.05$.

Additionally, a further selection of significant gene expression changes associated with the two *VRK2* siRNA conditions that occurred a nominally significant $P < 0.05$ and in the same direction (i.e. up- or down-regulation) relative to the negative control siRNA condition was performed. This was done to refine the data set to those changes reflecting *VRK2* knockdown rather than off-target effects of individual *VRK2* siRNA. The overlapping gene set resulting from this selection was subjected to Gene Ontology analysis (GO) through the DAVID Bioinformatics Resource 6.7 (Huang da et al., 2009), using all gene probes that had a detection *P* value > 0.95 in at least all four samples of one of the comparison conditions as the background comparison.

5.3 Results

5.3.1 Transfection efficiency of siRNA and DNA vectors.

To determine transfection efficiency, cells transfected with the BLOCK-iTTM Alexa Fluor® Red Fluorescent oligo or the pmaxGFP® Vector were checked 24 hours post-transfection by fluorescence microscopy. Transfection efficiency was very high with >80% of cells presenting red or green fluorescence 24 hours post-transfection (Figure 5.6)

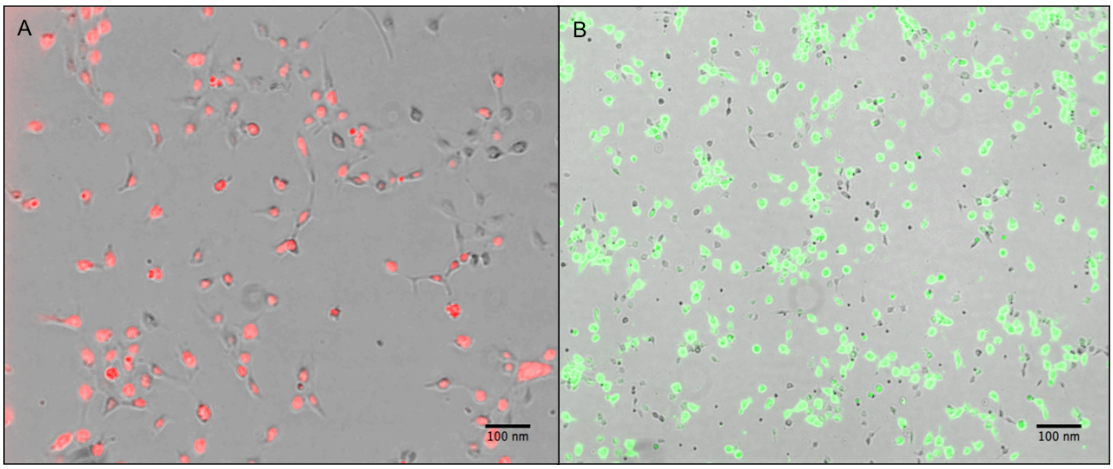


Figure 5.6 Assay of transfection efficiency using the NucleofectionTM technology from Lonza

Transfection efficiency was measured by the uptake of BLOCK-iT Alexa Fluor red Fluorescent Oligo (A) or pmaxGFP® Vector (B) into neuronal progenitor cells CTX0E16. Fluorescence was checked 24 hours post-transfection. Images were taken using the inverted microscope Olympus Ix70 using AxioVision 4.8 Imaging software (Zeiss). Scale bar 100 nm.

5.3.2 Knockdown of *VRK2* mRNA in CTX0E16 cells.

Confirmation of *VRK2* mRNA knockdown, relative to the negative control siRNA, was assessed by quantitative qPCR. The qPCR values for *VRK2* mRNA expression level were normalized against the geometric average of three internal control genes (*RPLI3A*, *GADPH* and *CFDP1)*. Table 5.7 show the normalized data for CTX0E16 cell line, used in the statistical analysis.

Table 5.7 Normalized *VRK2* expression values after RNAi transfection, obtained in CTX0E16 cell line. These values were used in statistical analysis applying a linear mixed model.

siRNA	Normalized values for *VRK2*				Biological replicate
	qPCR1	qPCR2	qPCR3	qPCR4	
siRNA1	0.4158	0.3706	0.3249	0.3688	1
siRNA1	1.1820	0.8482	0.7298	0.8805	2
siRNA1	0.9056	0.6906	0.6694	0.6026	3
siRNA1	0.5647	0.3781	0.2782	0.2900	4
siRNA2	1.6384	0.4008	0.2349	0.3621	1
siRNA2	0.4757	0.4344	0.4628	0.5074	2
siRNA2	0.4583	0.4281	0.4541	0.5112	3
siRNA2	0.6768	0.4889	0.4176	0.5729	4
control	2.0984	0.6679	0.6572	0.5794	1
control	0.2573	0.8983	0.8844	1.2668	2
control	1.4906	1.0181	1.0341	1.0036	3
control	1.2838	0.9038	0.6599	0.6045	4

Statistical analyses were performed applying a linear mixed model in order to identify expression changes in *VRK2* mRNA level between control siRNA and *VRK2* siRNA conditions. The results obtained from the statistical analysis are show in Table 5.8.

Table 5.8 Statistical result for *VRK2* expression after RNAi transfection, in the CTX0E16 cell line.

Treatment	Estimate	SE	Pr(>|t|)	Mean expression	mean-1SE	mean+1SE	Fold change (%)
Control	-0.1434	0.1162		0.8664	0.7714	0.9732	100
siRNA1	-0.6169	0.1162	0.0040**	0.5396	0.4805	0.6061	62.2819
siRNA2	-0.721	0.1162	0.0006 ***	0.4863	0.4330	0.5462	56.1244

VRK2 expression was represented relative to the negative control, Figure 5.7 show that siRNA1 and siRNA2 reduced *VRK2* expression by 38% and 44% respectively relative to control siRNA1, being both significant (P = 0.0040 and P = 0.0006, respectively).

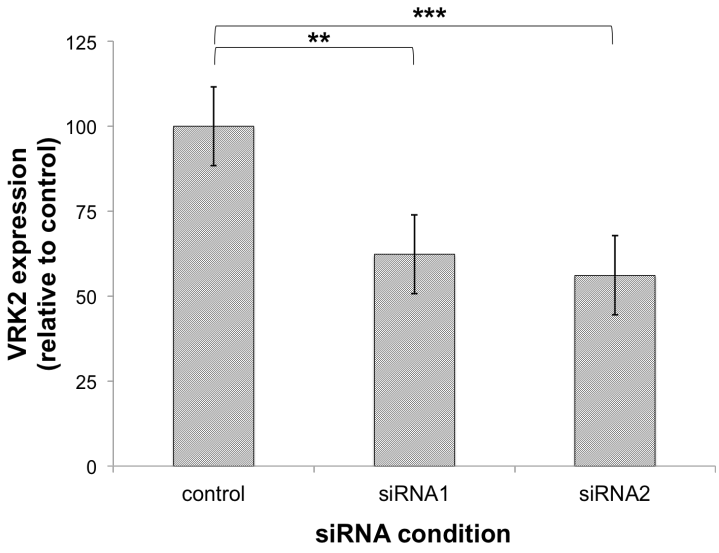


Figure 5.7 Expression of *VRK2* mRNA after RNAi transfection

QPCR results showing *VRK2* mRNA expression 120 hours post-transfection with two independent *VRK2* siRNAs (siRNA1 and siRNA2). Data are normalized against the geometric average of 3 housekeeper genes and plotted relative to the negative control siRNA. *VRK2* mRNA was reduced by 38 % and 44% in association with siRNA1 (P = 0.004) and siRNA 2 (P = 0.0006) respectively. Error bars represent SEM derived from four separate transfection experiments. Asterisks show where the expression of *VRK2* is significantly different from control. Asterisks show where the expression of *VRK2* is significantly different from control (P < 0.05).

5.3.3 Overexpression of *VRK2* mRNA in CTX0E16 cells.

Confirmation of *VRK2* mRNA overexpression, relative to the negative control pmaxGFP was assessed by quantitative qPCR. The qPCR values for *VRK2* mRNA expression level were normalized against the geometric average of three internal control genes (*RPLI3A*, *GADPH* and *CFDP1)*. Table 5.9 show the normalized data for CTX0E16 cell line, used in the statistical analysis.

Table 5.9 Normalized *VRK2* expression values after overexpression, obtained in CTX0E16 cell line. These values were used in statistical analysis applying a linear mixed model.

DNA vector	Normalized values for *VRK2*				Biological replicate
	qPCR1	qPCR2	qPCR3	qPCR4	
pVRK21	1.064911298	0.759224598	0.485146328	1.07645406	1
pVRK22	1.074722625	2.156801469	1.843783161	1.110078615	2
pVRK23	1.330978842	0.650571805	0.591582528	0.739045715	3
pVRK24	1.001125166	0.64745185	0.697898028	0.664713035	4
pGFP1	1.597601312	0.818814737	0.79111003	1.039437635	1
pGFP2	0.777440249	0.757331832	0.79898748	1.047042158	2
pGFP3	0.702510167	0.900261514	0.861260649	1.088427272	3
pGFP4	0.420753151	0.553121785	0.592258908	0.742052644	4

Statistical analyses were performed applying a linear mixed model in order to identify expression changes in *VRK2* mRNA level between control pmaxGFP and pVRK2 conditions. The results obtained from the statistical analysis are show in Table 5.10.

Table 5.10 Statistical result for *VRK2* expression after overexpression, in the CTX0E16 cell line.

Treatment	Estimate	SE	Pr(>|t|)	Mean expression	mean-1SE	mean+1SE	Fold change (%)
pGFP	-0.2149	0.1256		0.8066	0.7115	0.9146	100
pVRK2	-0.0924	0.1256	0.288	0.9117	0.8041	1.0337	113.0297

VRK2 expression was represented relative to the negative control. Transfection of CTX0E16 with pVRK2 produced an increase in the relative expression of VRK2 mRNA of 13%. However, this overexpression of *VRK2* mRNA was not significant relative to the control pmaxGFP (*P* = 0.288) (Figure 5.8)

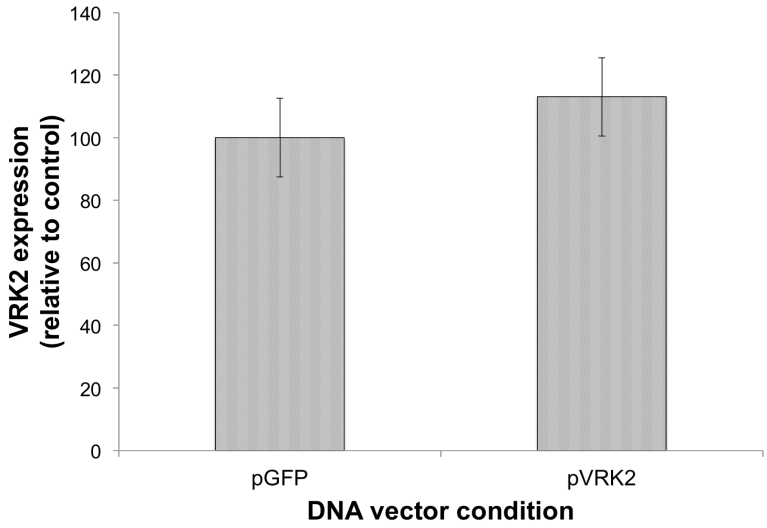


Figure 5.8 expression of *VRK2* mRNA after overexpression with pVRK2

QPCR results showing *VRK2* mRNA expression 120 hours post-transfection with pVRK2 and the negative control pmaxGFP. Data are normalized against the geometric average of 3 housekeeper genes and plotted relative to the negative control. A 13 % increase in VRK2 mRNA expression was observed in association with pVRK2, but this was not significant (*P* = 0.288). Error bars represent SEM derived from four separate transfection experiments.

5.3.4 Effect of *VRK2* knockdown on VRK2 protein in CTX0E16 cells.

Reduced levels of VRK2 protein associated with the siRNA conditions were verified by inmunobloting (Figure 5.9). VRK2 protein levels were normalized against the geometric average of three internal control proteins (α-tubulin, β-actin and, GADPH*)* and were represented relative to the negative control.

Knockdown of *VRK2* with siRNA1 and siRNA2 resulted in a reduction of 46 % and 48 % in VRK2 protein levels respectively.

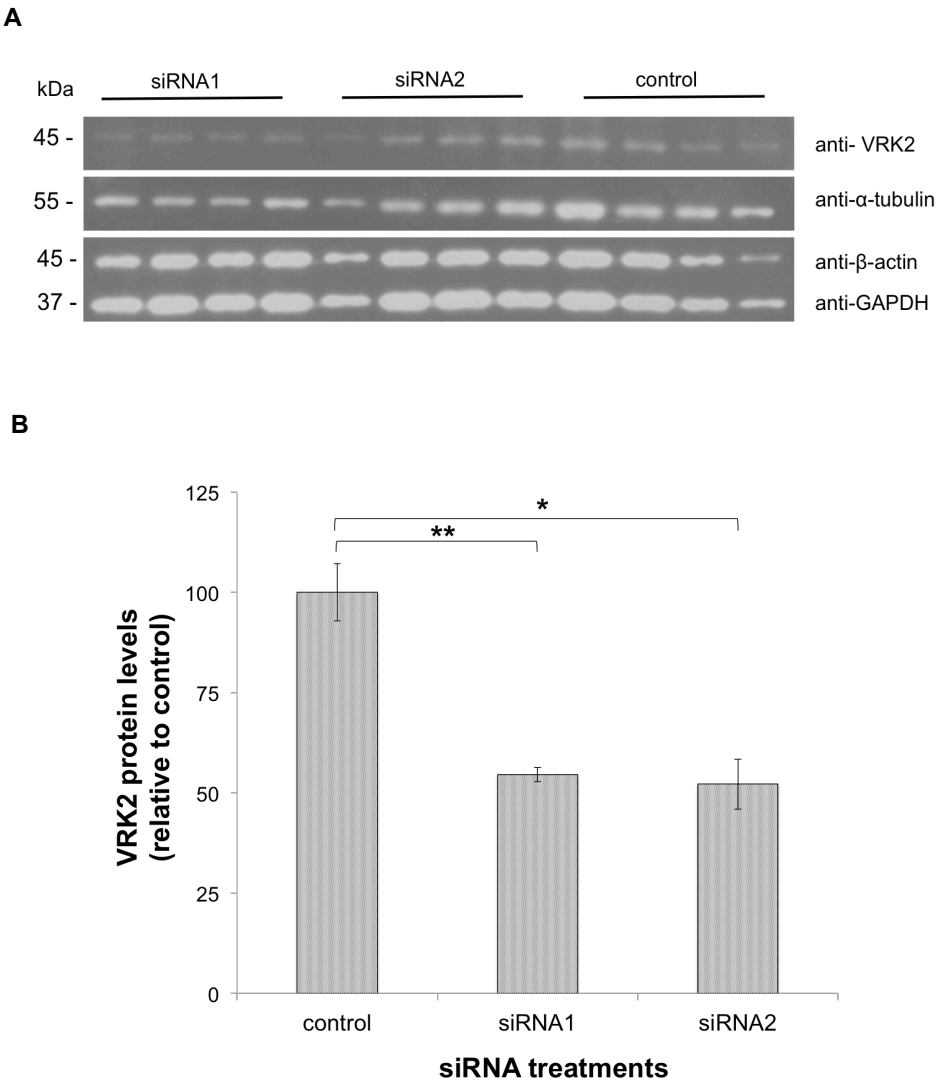


Figure 5.9 Expression of VRK2 protein after RNAi transfection

(A) Immunoblots for proteins lysates obtained from cells transfected with siRNA1, siRNA2 and Stealth siRNA™ negative universal control show that both siRNA1 and siRNA2 reduced VRK2 levels. The image was analyzed by densitometry using ImageJ. (B) VRK2 levels were normalized against the geometric average of 3 housekeeper proteins (α-tubulin, β-actin and, GADPH*)* and plotted relative to the negative control siRNA. VRK2 protein was reduced by 46 % and 48% in association with siRNA1 (*P* = 0.005) and siRNA2 (*P* = 0.05) respectively. Error bars represent SEM derived from four separate transfection experiments. Asterisks show where the expression of VRK2 is significantly different from control.

5.3.5 Proliferation of CTX0E16 cells.

The effect of VRK2 knockdown on the proliferation of CTX0E16 was analysed by ki67 inmunostaining. Background staining for ki67 was removed by quantifying positive ki67 cells from a negative control (Figure 5.10).

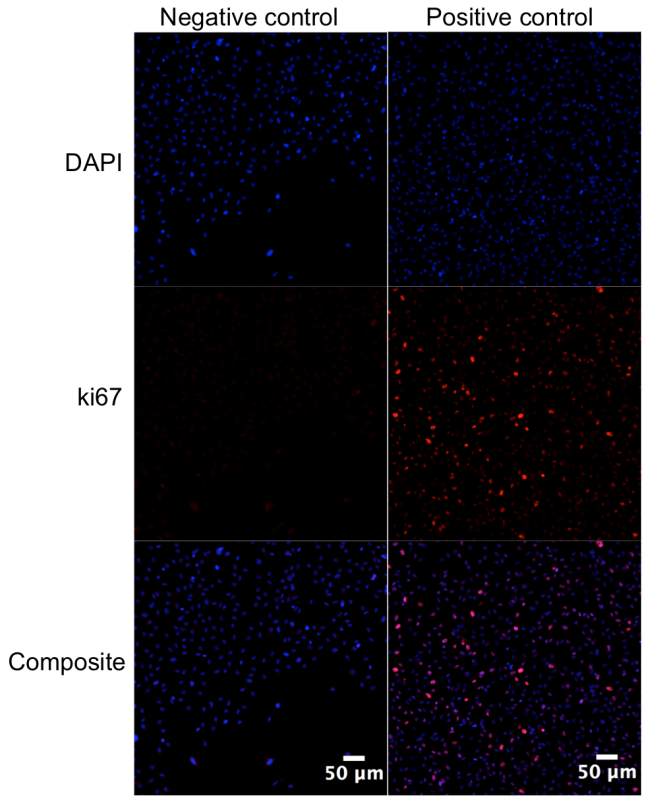


Figure 5.10 Examination of ki67 inmunostaining.

120 hours after transfection cell proliferation was evaluated by ki67 inmunostaining assay. For all conditions a negative control, which was only stained for secondary antibody and DAPI, was used to eliminate background (false positive staining for ki67). Positive ki67 cells were quantified using Cell Insight. Images were taken using Cell Insight (Thermo). Scale bar 50 µm.

Cell counts for proliferating cells (ki67 positives) and total number of cells per well for each treatment in three biological replicates (3 plates) were used in the statistical analyses. I could not use the four biological replicates because one of them had very low cell count. This is a disadvantage of the Nucleofection^{TM} technology.

Because the number of cells observed per well in each treatment and in the three biological replicates was variable, I decided to analyse my data using a mixed-model to account for this systematic variation. In addition, I used a Poisson linear mixed-model in order to account explicitly for the total number of cells in the well when comparing the fraction of proliferating cells. This helps avoid potentially misleading results due to very different number of cells across treatments or biological replicates. For instance, simply using the fraction of proliferating cells but ignoring the total number of cells would deem a ratio of 50% equal from a well with 100 or 10000 total cells, when in the second case there is much more certainty that the ratio is 50%. The analysis considers treatments as fixed effects and aims to account for the systematic variation between by modelling biological replicates as random effects. Below, I will describe in detail the process used to summarise and filter the data in order to perform statistical analyses as well as checks I did to understand the statistical results and their robustness.

Data analysis

1. Firstly, I explored the variability in total number of cells per biological replicate and well (Figure 5.11, top). The three plates had low cell counts in all the treatments, with only plate 1 having higher number of cells in the control condition. I also examined the fraction of proliferative cells (Figure 5.11, bottom); plate 1 showed the greatest variability in the raw proliferation ratio.

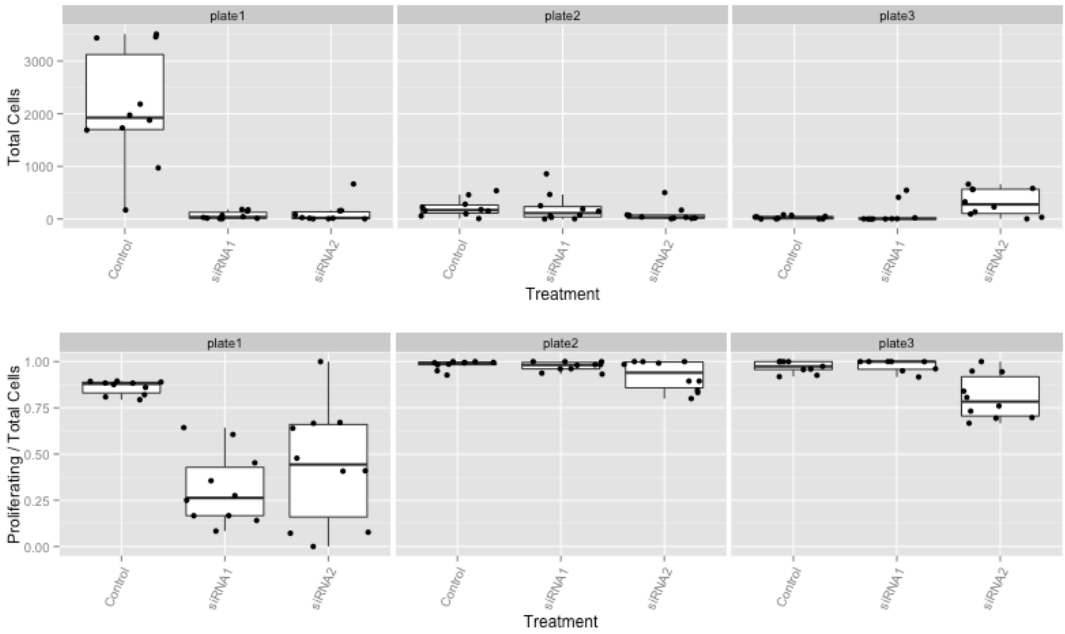


Figure 5.11 Total number of cells for each treatment per plate (top) and fraction of proliferative cells for each treatment per plate (bottom), for CTX0E16 cell line.

2. Then I examined the relationship between total number of cells and number of proliferating cells (Figure 5.12). This confirmed previous observations for plate 1 in the control condition and showed that in general wells have a very low number of proliferating cells, which coincide with the observe low cell counts. It also showed that plate 3 control condition has the lowest number of cell than any other plate or condition (all wells are on the bottom left corner).

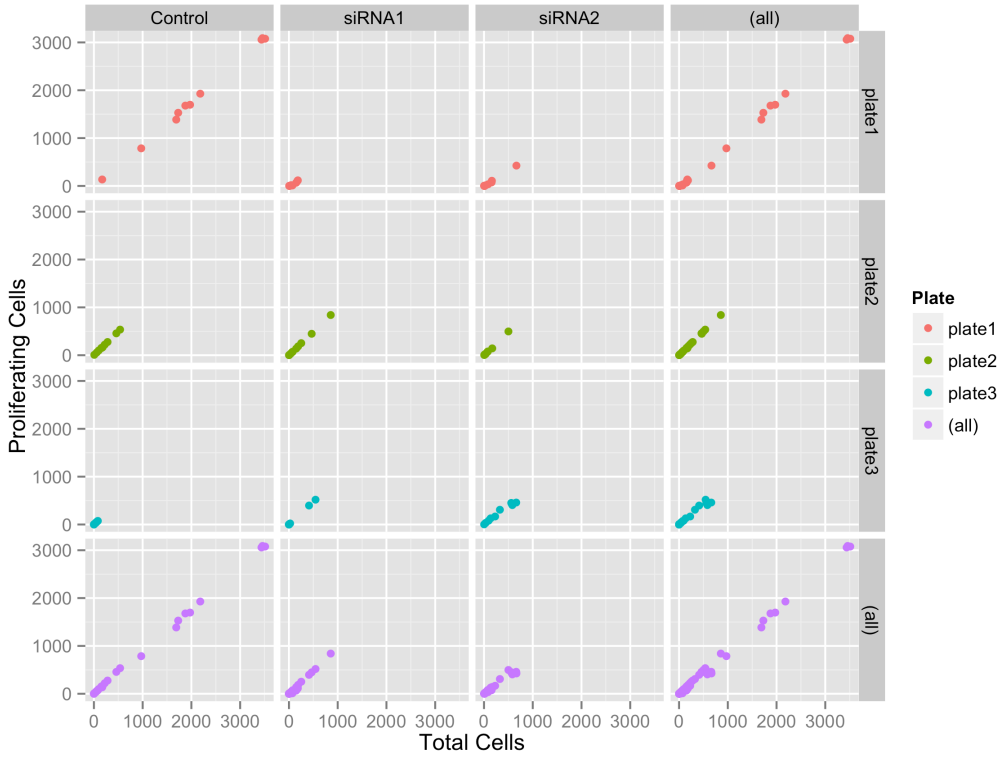


Figure 5.12 Relationship between total number of cells and proliferating cells for each treatment per plate, observed on CTX0E16 cell line

3. Based on observations from point 2, I decided to remove from the analysis wells with less than 100 cells and to remove plate 3 due to the low cell counts observed in the control siRNA treatment. This leads to the following dataset: Figure 5.13.

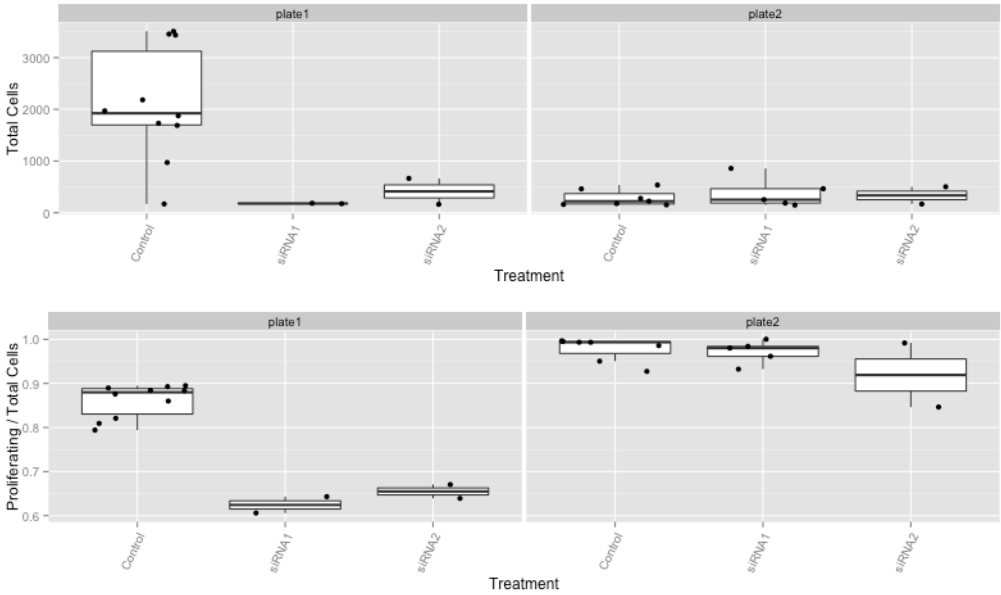


Figure 5.13 Total number of cells for each treatment per plate (top) and fraction of proliferative cells (bottom). These data excluded plate 3 and wells with less than 100 cells. Data obtained on CTX0E16 cell line.

4. In order to take into account the difference in the total number of proliferating cells in different wells I used a Poisson model, which explicitly models this. As explained before, using the ratio of proliferating/total number of cells would ignore differences in total number of cells and regards as equal a case with ratio of 50% from counts 10/20 or 1000/2000 in spite that the second case has a much more certainty that the ratio is 50%. In order to account for the systematic variation across wells and biological replicates I used a generalized linear mixed model.

    a) First I analysed the data removing plate 3 and wells with less than 100 proliferating cells. The statistical results are show in Table 5.11.

Table 5.11 Statistical results for proliferating CTX0E16 cells with plate 3 removed and removed wells with less than 100 proliferating cells.

Treatment	Estimate	SE	Pr(>|z|)	Mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.0508	0.0646		0.9504	0.8910	1.0138	100
siRNA1	-0.1470	0.0677	0.0005 ***	0.8633	0.8068	0.9238	90.8346
siRNA2	-0.2380	0.0701	1.67e-09 ***	0.7882	0.7349	0.8454	82.9353

b) Finally, in order to compare I re-analysed the data without filtering, the statistical results are show in Table 5.12.

Table 5.12 Statistical result for proliferating CTX0E03 cells, without filtering.

Treatment	Estimate	SE	Pr(>|z|)	Mean proliferation	mean-1SE	mean+1SE	Fold change (%)
Control	-0.0102	0.059		0.9899	0.9331	1.0500	100
siRNA1	-0.1708	0.0603	8.53e-13 ***	0.8430	0.7937	0.8954	85.1633
siRNA2	-0.2862	0.0601	< 2e-16 ***	0.7511	0.7073	0.7976	75.8813

The comparison of both statistical results suggests that both results are consistent (see mean estimates and their standard errors). However, because cell counts in the control condition of plate 3 were very low I decided to used the filtered data for >100 cells for discussion.

Summary of proliferation results for CTX0E16 cell line:

Proliferative cells were reduced when *VRK2* was knocked down with siRNA1 and siRNA2, relative to the negative control siRNA condition. Table 5.11 show that a 9.2% reduction in cell proliferation occurs with siRNA1 and 17% with siRNA2 treatment. These two results are statistically significant, P =0.0005 and $P < 1.67 \times 10^{-9}$, respectively.

5.3.6 Microarray analysis of VRK2 knockdown

A total of 23657 probes were detected with detection *P* > 0.95 in at least all four cellular RNA samples derived from one of the comparison conditions (siRNA1, siRNA2 or negative control siRNA). These probes correspond to transcripts for 16047 unique Refseq genes, of which 16001 mapped to DAVID IDs.

We did not find evidence of differential expression for any probe at FDR < 0.05. However, using a nominal significance threshold (*P* < 0.05) gene expression differed between siRNA1 and the negative control siRNA condition at 760 probes, and expression at 1104 probes differed (*P* < 0.05) between siRNA2 and the negative control siRNA condition. Gene expression differing between both VRK2 siRNA conditions and occurring in the same direction (i.e. up- or down-regulation) relative to the negative control siRNA condition was observed at 186 probes. By chance alone we expected 14.7 probes to overlap in direction of effect at *P* < 0.05, this comes from each probe having 5% chance of being on the list with a *P* < 0,05 and 50% chance to have either direction of effects and, therefore, the probability of finding the same probe on the list for both siRNA is 0.05*0.05*0.5*0.5 = 0.000625. Thus, the overlap of the two lists is ~12 (or 186/14.7) higher than expected by chance. Of these, 73 were up-regulated and 114 were down-regulated, in association with the two *VRK2* siRNA conditions. These probes are listed in Table 5.13.

Table 5.13 Genes showing differential expression (*P* < 0.05), in the same direction relative to the negative control siRNA condition, in association with both siRNA *VRK2* conditions (siRNA1 and siRNA2).

Gene Symbol	pvalue_siRNA1	pvalue_siRNA2	coefficient_siRNA1	coefficient_siRNA2
Up-regulated genes				
TDGF1	0.000562502	0.015993001	0.108894738	0.065746831
PARVB	0.001085866	0.000112496	0.162622884	0.213724397
AI445550	0.001365606	0.006555432	0.103773374	0.082286977
RMDN1	0.002942841	0.02696716	0.182234814	0.123614249
PRPSAP2	0.003034752	0.011569297	0.190162548	0.15304456
C10orf32	0.004484996	0.024736302	0.188158312	0.138540757
DCUN1D3	0.004689574	0.011744062	0.142917893	0.122535533
RGNEF	0.004716534	0.020442091	0.10258376	0.079187238
TMEM14C	0.005739044	0.037346351	0.193072885	0.134851175
TRAPPC4	0.005756726	0.003204869	0.165322168	0.180963296
GPS1	0.006612141	0.036460833	0.178045339	0.127965522
HSFY1, HSFY2	0.006954177	0.003455626	0.097242572	0.108547943
LILRA4	0.007198439	0.049813285	0.094235205	0.063649731
ANKRD20A4	0.007823151	0.019291368	0.137449636	0.116525004
HIST1H4C	0.008874388	0.000404771	0.378797227	0.58700414
SPATA24	0.009959676	0.032172959	0.108666208	0.086161711
CAP1	0.010002136	0.00211621	0.225879075	0.288082019
LOC100129233	0.01000256	0.044690056	0.109945417	0.080729547
TIMM10	0.012271652	0.0042838	0.178035058	0.212272145
SATB1	0.012632987	0.018565852	0.112724042	0.10475354
URM1	0.01308214	0.003544943	0.109511733	0.135975118
AP4B1	0.013347071	0.025500317	0.135568207	0.1192396
TDG	0.013693775	0.002797473	0.398735803	0.517123665
TRIM69	0.014198396	0.003630487	0.145446017	0.182707559
FKBP14	0.016731267	0.03542051	0.286738231	0.244726163
NEK8	0.017199312	0.008360029	0.132044749	0.150564639

NME2	0.018237065	0.004598015	0.089223862	0.113429059
BOLA2B	0.018299326	0.004733067	0.111064749	0.140653541
LOC647323	0.018490338	0.038378472	0.080960241	0.069153874
CCL7	0.019271486	0.025282282	0.07465258	0.070594249
PPP1R14B	0.020599754	0.035806463	0.158521788	0.140634181
CHCHD1	0.022042898	0.000227044	0.08754136	0.172347624
SIVA1	0.023637116	0.013237825	0.196719159	0.220441449
WWOX	0.023696438	0.002296837	0.13387213	0.199020034
DERL2	0.024235447	0.005219824	0.144014721	0.190172688
LEMD3	0.024401448	0.006189997	0.128359244	0.165162351
MRPL19	0.024778145	0.001833697	0.154065338	0.238648202
IKZF3	0.025947828	0.027317072	0.104234656	0.10308801
GTPBP8	0.02631045	0.000347372	0.133452197	0.259325751
RABL6	0.027108816	0.014815161	0.086454005	0.097664599
CSH2	0.027950342	0.016341573	0.069668424	0.077756175
POLR3F	0.028428399	0.005522711	0.120188702	0.162766285
LOC613206	0.028560332	0.014744227	0.069946322	0.07999019
SLC35B1	0.029334987	0.00285043	0.137433148	0.207326404
PRO0611	0.029599822	0.033734001	0.070026468	0.068000092
KIAA2013	0.036807127	0.011417571	0.097642864	0.123965404
PVRL3	0.036909957	0.02612173	0.276359121	0.298584447
HAL	0.037667718	0.015011244	0.07090034	0.086043894
PLEKHG7	0.03774026	0.018117348	0.107179341	0.125485816
MYL12B	0.039583775	0.044708841	0.143617241	0.139430759
IGHG1	0.040730468	0.013665639	0.06263689	0.078821485
ITGAE	0.042050026	0.018237144	0.182154977	0.218571311
SNORA27	0.043540208	0.01935361	0.101172768	0.120999481
TMEM189-UBE2V1	0.043556741	0.048302016	0.089758284	0.087474135
TPSAB1	0.043738886	0.046670055	0.06964183	0.068529908
ATM	0.043874994	0.04220882	0.07046269	0.071133044
TDP1	0.044387372	0.021259967	0.227938905	0.268755905
AGO2	0.044749699	0.032835495	0.226977968	0.244239789

SERPINB2	0.045236381	0.03572296	0.06186777	0.065473266
ZMAT3	0.045804711	0.004634571	0.128419265	0.199814865
COX11	0.046143623	0.018862172	0.063641551	0.077612275
C7orf55	0.046602439	0.005850509	0.200221528	0.301542355
TBX20	0.046848155	0.01969235	0.060775189	0.073761108
RS1	0.047575471	0.048841409	0.094497173	0.093871283
FAM106A	0.048036891	0.04634262	0.070360139	0.070998151
Down-regulated genes				
RNFT2	0.000780343	0.012574238	-0.157913896	-0.103945399
DCLK1	0.001001211	0.005699693	-0.290390107	-0.226038296
CYP2D6	0.001517561	0.040931764	-0.115474136	-0.064860085
IPCEF1	0.001527876	0.002251911	-0.161005181	-0.152539063
DOCK1	0.001724728	0.034381134	-0.132513671	-0.078921298
POLR1A	0.001836911	0.025089912	-0.120452769	-0.077574601
ST6GAL1	0.001954262	0.001435882	-0.229899837	-0.239869289
DGCR2	0.002354671	0.000425136	-0.160599681	-0.201083166
EPHA6	0.002604968	0.02128854	-0.102870335	-0.072036574
ATP1A1	0.002786111	0.004968869	-0.250349693	-0.229354253
AHNAK	0.002978134	0.003731833	-0.272209645	-0.263201163
EIF4B	0.003763348	0.001579511	-0.265284115	-0.300554233
PSG1	0.004060261	0.045760397	-0.112373629	-0.070778042
CYP39A1	0.004679973	0.017896486	-0.123924834	-0.098112284
TAOK2	0.004852621	0.003176935	-0.126441692	-0.134875116
AMZ2	0.006440712	0.002123481	-0.130856495	-0.154824546
ZFYVE26	0.00702917	0.007859664	-0.14551965	-0.142827269
ANKHD1-EIF4EBP3	0.007168478	0.012702969	-0.110277432	-0.099799163
AP3B1	0.008060022	0.038853745	-0.125190085	-0.091616673
TRIM13	0.008353012	0.013372396	-0.144087343	-0.132535292
DLL1	0.008800629	0.02985976	-0.128384098	-0.101306428
RERE	0.009283315	0.016415567	-0.19955757	-0.179811888
JMY	0.009485717	0.010700525	-0.211169714	-0.206739052
NHSL2	0.009601524	0.025984737	-0.091800687	-0.07579751

ZNF490	0.009702641	0.014468097	-0.165404619	-0.153848089
DYNLRB1	0.010160281	0.048396421	-0.206905124	-0.149334814
ZNF681	0.01022604	0.037503261	-0.105201316	-0.080895573
BM665736	0.01057738	0.005113351	-0.111751292	-0.126193002
LOC388692	0.011108609	0.002645896	-0.085119065	-0.107119267
RAB4A	0.011173148	0.004092901	-0.100201517	-0.118288295
ADH6	0.011474485	0.031267651	-0.102582795	-0.083948896
HOXA3	0.011553113	0.002247779	-0.088183234	-0.114408212
ZC3H11A	0.012118129	0.014695571	-0.154044295	-0.148637023
OSBPL2	0.012143725	0.010014164	-0.126358733	-0.130791698
ARHGEF4	0.012757679	0.00785094	-0.09856025	-0.107347009
GRIA1	0.013099727	0.045148706	-0.134247586	-0.10321889
DCTN1-AS1	0.013166385	0.049273455	-0.079656835	-0.059970598
OCRL	0.013167671	0.00438068	-0.138898402	-0.16718732
LOC100130480	0.013859029	0.0276356	-0.088939846	-0.077426848
ERCC6	0.014151489	0.005476056	-0.084841301	-0.099933671
APLF	0.015114441	0.00450578	-0.0942608	-0.115919875
TIMP3	0.015465339	0.031911647	-0.219456655	-0.188937054
RBL2	0.016244972	0.008717631	-0.12660388	-0.141753419
CDO1	0.01629827	0.047115425	-0.213282384	-0.169074942
DDX17	0.016638874	0.010175059	-0.178768864	-0.195754012
FMN1	0.017533046	0.034411144	-0.10310382	-0.089397159
EIF4B	0.017659595	0.000741158	-0.237100001	-0.386668222
TMED10	0.01769431	0.000579471	-0.221133802	-0.372132724
SH2D3C	0.018252657	0.011836221	-0.087792259	-0.095279385
NACAP1	0.018253896	0.026628278	-0.251248055	-0.232452602
SNORD59B	0.018382975	0.009357225	-0.08692449	-0.098491053
C1orf140	0.018437787	0.010483364	-0.085699431	-0.09524226
DVL3	0.018940265	0.032105339	-0.163158107	-0.145924608
EEF1G	0.01938727	0.011195302	-0.197917378	-0.219576873
MPG	0.019621542	0.017963816	-0.089208065	-0.09078551
HABP2	0.020351517	0.028854455	-0.100713797	-0.093579324

SNRPN	0.02061835	0.038170653	-0.148733828	-0.130013084
DPAGT1	0.021210168	0.049096484	-0.091890116	-0.075953408
TBC1D	0.021291713	0.00174244	-0.086742374	-0.131122775
PIGZ	0.021749144	0.00633156	-0.096135796	-0.120332482
TIGD1	0.022364502	0.002801791	-0.084567944	-0.120709416
OS9	0.022567551	0.011879717	-0.120496646	-0.136410244
FBXL8	0.022584228	0.033642304	-0.074872957	-0.068665526
TBL1XR1	0.023009772	0.010678283	-0.199998378	-0.231716767
ST3GAL1	0.02307969	0.030755188	-0.129256673	-0.121510513
PARP6	0.024004143	0.003697381	-0.084727252	-0.117777708
MBTPS1	0.024166971	0.046242439	-0.126390126	-0.108964303
MEGF8	0.025170328	0.011589213	-0.09100047	-0.105874544
TET3	0.026521608	0.011541223	-0.076797052	-0.090420603
PLEKHH1	0.02695187	0.02311532	-0.10937158	-0.112987466
GOLGA8B	0.027812793	0.017063978	-0.101281992	-0.111976159
CYP2U1	0.028019221	0.001306604	-0.101065132	-0.168350316
BF510292	0.028497946	0.006570487	-0.072408921	-0.095392021
PLBD2	0.029199887	0.025414546	-0.138316206	-0.142533869
PSG7	0.029593212	0.002992824	-0.066118783	-0.099225083
CDY1B	0.030745826	0.022470672	-0.103416915	-0.110613634
RFX2	0.030801713	0.042607595	-0.116428156	-0.107947124
SLC28A1	0.03114922	0.008392923	-0.087043761	-0.112260069
LPHN1	0.031361714	0.026934556	-0.177923603	-0.183977194
APLP2	0.031416411	0.022244413	-0.163533916	-0.17612763
LOC644172	0.03238532	0.029601484	-0.094192005	-0.096102509
LGI3	0.032400521	0.016442237	-0.075763651	-0.087272443
ZNF106	0.032445073	0.028857252	-0.238568249	-0.244877393
BTN3A3	0.032681786	0.040390835	-0.116110502	-0.110515759
SDHAP1	0.033250454	0.035993594	-0.128867577	-0.126538966
GSAP	0.03357984	0.038698394	-0.102397626	-0.099073192
PAIP1	0.035548359	0.036436877	-0.075008563	-0.074579264
COL4A5	0.035641767	0.011027767	-0.242370469	-0.307320414

HAGH	0.036644743	0.024250158	-0.100710863	-0.11035424
TNFRSF19	0.037432769	0.048173187	-0.084300764	-0.079273494
ZNF623	0.037565632	0.041901728	-0.082424359	-0.080301
ZNF23	0.038268478	0.001656927	-0.106328404	-0.184058816
IRF2BP2	0.038436279	0.004937119	-0.126021494	-0.186057426
DNAH9	0.038846618	0.027933749	-0.066911643	-0.072118061
SLC26A3	0.040315742	0.007521603	-0.078384291	-0.109323331
CACNB2	0.041250155	0.011322676	-0.105734357	-0.138131497
PCBP2	0.041376541	0.042823288	-0.144939574	-0.143734178
KRT8P41	0.041412628	0.042083529	-0.077966545	-0.077663402
RBBP8NL	0.042499429	0.011468001	-0.065808294	-0.086398206
EFHD1	0.044697831	0.00492456	-0.139686986	-0.213988896
TMEM164	0.045902708	0.019178661	-0.098079302	-0.119058607
PDGFD	0.046284602	0.03473703	-0.177677981	-0.190335857
ZNF580	0.047555963	0.047334681	-0.085923715	-0.08602469
WNK2	0.047887172	0.011865333	-0.112249787	-0.150867267
ACADSB	0.048560625	0.027199283	-0.104304442	-0.119436325
TRIM33	0.048679622	0.037289537	-0.138696578	-0.148016893
SNX30	0.049283684	0.00360763	-0.108899394	-0.179349825
EGFEM1P	0.049904746	0.04777355	-0.141701862	-0.143281808

Note: Coefficient indicates differential expression value.

Differentially expressed genes were tested for enrichment in particular GO terms using the bioinformatic online database DAVID. The most significant terms for differentially expressed genes compared with the background list were observed on the cellular component process of 'microsome' and 'vesicular fraction' (GOTERM_CC_FAT) and in the subcategories relating to 'intracellular part', 'intracellular' and 'vesicular fraction' (Bonferroni corrected $P < 0.05$). The 10 differentially expressed genes belonging to this GOTERM_CC_FAT process are *RAB4A*, *TMED10*, *CYP2D6*, *ATP1A1*, *CYP2U1*, *DGCR2*, *SLC35B1*, *DPAGT1* and *CYP39A1*. Additional GO terms in this category relating

to ‘cell fraction’, ‘organelle membrane’ and ‘vesicles’ were also significantly enriched among differentially expressed genes, although, they did not survive Bonferroni correction for all terms in the category.

Compared with the background list, genes annotated as belonging to the biological process (GOTERM_BP) of ‘cellular response to stress’, ‘stress activated protein kinase and ‘JNK cascade’ were statistically overrepresented in this category (P < 0.05 uncorrected). A significant enrichment was observed for the molecular function term (GOTERM_MF) of ‘nucleic acid binding’ and ‘DNA binding’. However, for both GO categories the significance did not survive conservative Bonferroni corrections for all terms in each category. All significant GO terms for differentially expressed genes are shown in Table 5.14. There was no significant enrichment for differentially expressed genes in any KEGG pathway.

When up-regulated genes and down-regulated genes were analyzed separately additional GO terms were significantly overrepresented for each set of genes. The most significant GO terms for up-regulated genes were ‘intracellular part’ in the cellular component category (Bonferroni corrected P < 0.0128). Likewise, genes annotated as belonging to the biological process term (GOTERM_BP) of ‘regulation of cell death’, ‘negative regulation of biological process’ and ‘cellular response to stimulus’ among others were statistically overrepresented in this set of genes. Additionally, a significant enrichment of genes belonging to the terms ‘nucleus’ and ‘intracellular membrane-bounded organelle’ from the categories GOTERM_CC_4 and GOTERM_CC_5 respectively, as well as from the category GOTERM_MF of ‘double-stranded DNA binding’ was observed in up-regulated genes (P < 0.05 uncorrected).

When down-regulated genes were analysed separately a significant enrichment was observed for ‘endoplasmic reticulum membrane’, ‘microsome’, ‘endomembrane system’, and ‘cytoplasmic membrane-bounded vesicle’ in the cellular component category (GOTERM_CC) together with ‘electron carrier activity’ and ‘oxidoreductase activity’ in the

molecular function category (GOTERM_MF) (P < 0.05 uncorrected). Figure 5.14 represent the GO terms enriched in down-regulated, up-regulated and the overlapping set of differentially expressed genes compared with the background list.

Table 5.14 GO functional enrichment analysis of differentially expressed genes after VRK2 knockdown in two independent siRNA conditions

Category	Term	Count	PValue	Bonferroni
GOTERM_CC_4	GO:0044424~intracellular part	101	0.0004	0.0053
GOTERM_CC_2	GO:0005622~intracellular	104	0.0003	0.0104
GOTERM_CC_FAT	GO:0005792~microsome	10	0.0001	0.0114
GOTERM_CC_FAT	GO:0042598~vesicular fraction	10	0.0001	0.0127
GOTERM_CC_3	GO:0005622~intracellular	104	0.0001	0.0153
GOTERM_CC_5	GO:0042598~vesicular fraction	10	0.0001	0.0155
GOTERM_CC_2	GO:0044424~intracellular part	101	0.0006	0.0195
GOTERM_CC_3	GO:0044424~intracellular part	101	0.0003	0.0312
GOTERM_CC_2	GO:0000267~cell fraction	19	0.0024	0.0800
GOTERM_CC_3	GO:0000267~cell fraction	19	0.0022	0.2042
GOTERM_CC_FAT	GO:0000267~cell fraction	19	0.0012	0.2165
GOTERM_CC_FAT	GO:0005624~membrane fraction	16	0.0012	0.2170
GOTERM_CC_4	GO:0005624~membrane fraction	16	0.0017	0.2314
GOTERM_CC_1	GO:0043226~organelle	82	0.0242	0.2364
GOTERM_CC_3	GO:0005626~insoluble fraction	16	0.0028	0.2576
GOTERM_CC_5	GO:0005624~membrane fraction	16	0.0020	0.2900
GOTERM_CC_FAT	GO:0005626~insoluble fraction	16	0.0017	0.2969
GOTERM_CC_1	GO:0044422~organelle part	44	0.0322	0.3021
GOTERM_CC_4	GO:0005626~insoluble fraction	16	0.0025	0.3130
GOTERM_CC_2	GO:0043227~membrane-bounded organelle	76	0.0215	0.5320
GOTERM_CC_2	GO:0012505~endomembrane system	13	0.0235	0.5653
GOTERM_MF_2	GO:0003676~nucleic acid binding	36	0.0276	0.6343
GOTERM_CC_3	GO:0044431~Golgi apparatus part	8	0.0109	0.6842
PANTHER_BP	BP00036: DNA repair	6	0.0104	0.7163
GOTERM_CC_4	GO:0043231~intracellular membrane-bounded organelle	76	0.0089	0.7435
GOTERM_CC_2	GO:0043229~intracellular organelle	82	0.0389	0.75087
GOTERM_CC_2	GO:0044446~intracellular organelle part	44	0.0396	0.7572
GOTERM_CC_2	GO:0031090~organelle membrane	15	0.0406	0.7656
GOTERM_CC_2	GO:0031982~vesicle	11	0.0407	0.7669
GOTERM_CC_2	GO:0044422~organelle part	44	0.0426	0.7823
GOTERM_CC_4	GO:0044431~Golgi apparatus part	8	0.0102	0.7892
GOTERM_CC_3	GO:0043231~intracellular membrane-bounded organelle	76	0.0161	0.8186
GOTERM_CC_FAT	GO:0044431~Golgi apparatus part	8	0.0085	0.8324
GOTERM_CC_5	GO:0044431~Golgi apparatus part	8	0.0111	0.8475
GOTERM_CC_3	GO:0044432~endoplasmic reticulum part	8	0.0218	0.9011
GOTERM_CC_3	GO:0012505~endomembrane system	13	0.0222	0.9057
GOTERM_CC_4	GO:0043229~intracellular organelle	82	0.0161	0.9151
GOTERM_CC_5	GO:0043231~intracellular membrane-bounded organelle	76	0.0147	0.9186
GOTERM_CC_4	GO:0044432~endoplasmic reticulum part	8	0.0204	0.9566
GOTERM_CC_3	GO:0043229~intracellular organelle	82	0.0300	0.9590

GOTERM_CC_FAT	GO:0012505~endomembrane system	13	0.0155	0.9626
GOTERM_CC_3	GO:0031410~cytoplasmic vesicle	11	0.0311	0.9636
GOTERM_CC_FAT	GO:0044432~endoplasmic reticulum part	8	0.0172	0.9737
GOTERM_CC_3	GO:0044446~intracellular organelle part	44	0.0348	0.9757
GOTERM_CC_5	GO:0044432~endoplasmic reticulum part	8	0.0221	0.9770
PANTHER_BP	BP00061: Protein biosynthesis	8	0.0313	0.9780
GOTERM_CC_3	GO:0031300~intrinsic to organelle membrane	5	0.0369	0.9807
GOTERM_CC_3	GO:0031090~organelle membrane	15	0.0383	0.9835
GOTERM_CC_4	GO:0044446~intracellular organelle part	44	0.0271	0.9846
GOTERM_CC_4	GO:0031410~cytoplasmic vesicle	11	0.0287	0.9880
GOTERM_CC_5	GO:0043229~intracellular organelle	82	0.0273	0.9906
GOTERM_CC_FAT	GO:0031410~cytoplasmic vesicle	11	0.0230	0.9925
GOTERM_CC_4	GO:0044444~cytoplasmic part	49	0.0327	0.9936
GOTERM_CC_4	GO:0031090~organelle membrane	15	0.0347	0.9953
PANTHER_MF	MF00124:Oxygenase	4	0.0439	0.9954
PANTHER_MF	MF00071: Translation factor	4	0.0439	0.9954
GOTERM_CC_5	GO:0031410~cytoplasmic vesicle	11	0.0315	0.9955
GOTERM_CC_4	GO:0031300~intrinsic to organelle membrane	5	0.0354	0.9959
GOTERM_CC_FAT	GO:0031090~organelle membrane	15	0.0263	0.9963
GOTERM_CC_4	GO:0005795~Golgi stack	3	0.0365	0.9965
GOTERM_CC_5	GO:0044446~intracellular organelle part	44	0.0351	0.9976
PANTHER_BP	BP00202:Other developmental process	4	0.0489	0.9976
GOTERM_CC_FAT	GO:0031982~vesicle	11	0.0291	0.9980
GOTERM_BP_4	GO:0031098~stress-activated protein kinase signalling pathway	4	0.0135	0.9982
GOTERM_BP_5	GO:0007254~JNK cascade	4	0.0126	0.9984
GOTERM_CC_5	GO:0031300~intrinsic to organelle membrane	5	0.0373	0.9984
GOTERM_CC_5	GO:0005795~Golgi stack	3	0.0376	0.9985
GOTERM_CC_FAT	GO:0031300~intrinsic to organelle membrane	5	0.0319	0.9989
GOTERM_BP_5	GO:0031098~stress-activated protein kinase signalling pathway	4	0.0136	0.9990
GOTERM_CC_FAT	GO:0005795~Golgi stack	3	0.0343	0.9993
GOTERM_CC_5	GO:0044444~cytoplasmic part	49	0.0431	0.9994
GOTERM_BP_3	GO:0033554~cellular response to stress	11	0.0266	0.9997
GOTERM_MF_ALL	GO:0003676~nucleic acid binding	36	0.0266	0.9999
GOTERM_BP_FAT	GO:0007254~JNK cascade	4	0.0129	1
GOTERM_BP_FAT	GO:0031098~stress-activated protein kinase signalling pathway	4	0.0139	1
GOTERM_BP_FAT	GO:0000165~MAPKKK cascade	6	0.0217	1
GOTERM_BP_FAT	GO:0033554~cellular response to stress	11	0.0305	1
GOTERM_BP_FAT	GO:0006284~base-excision repair	3	0.0367	1
GOTERM_CC_3	GO:0044444~cytoplasmic part	49	0.0429	1
GOTERM_MF_ALL	GO:0003677~DNA binding	27	0.0385	1

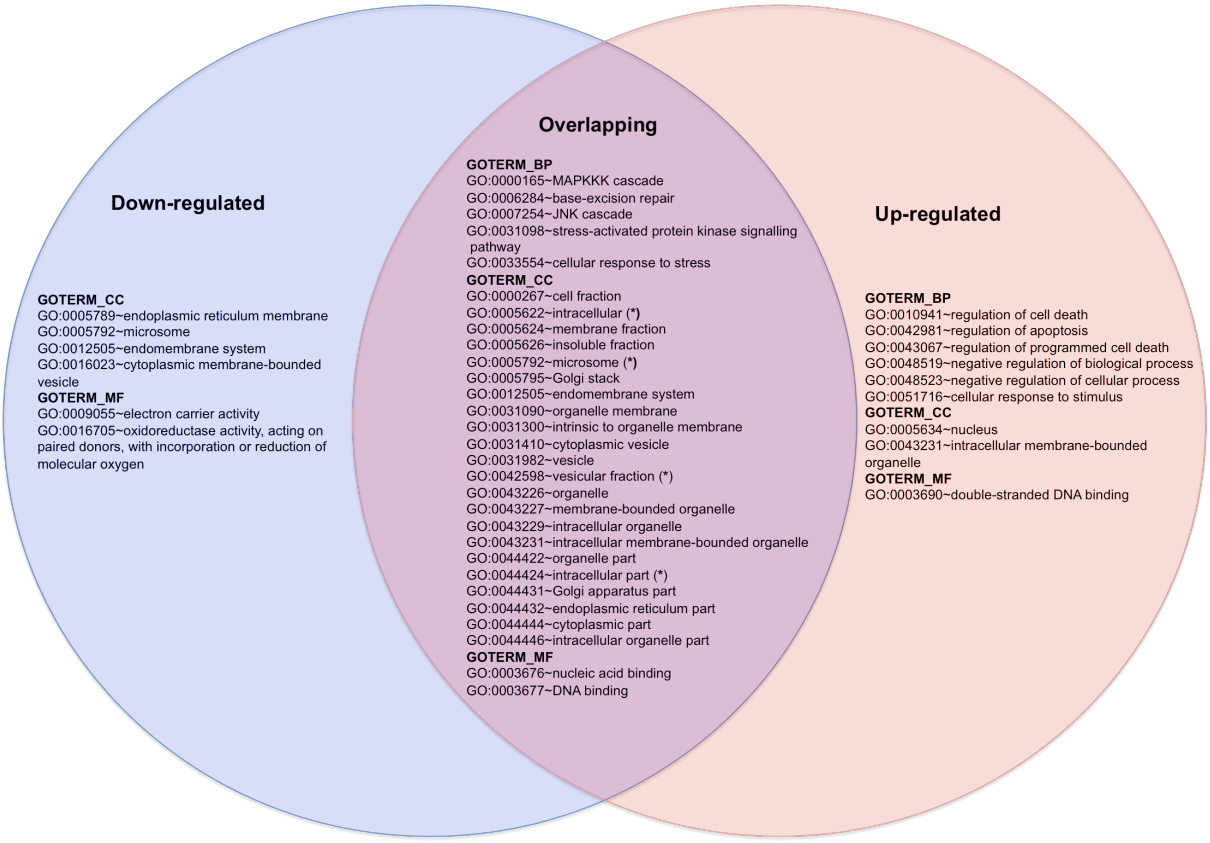


Figure 5.14 GO terms enriched in down-regulated, up-regulated and the overlapping set of differentially expressed genes compared with the background list. GO terms marked (*) survived Bonferroni corrections.

5.4 Discussion

The aim of the experiments presented in this chapter was to assess the effect of manipulating endogenous *VRK2* on the cellular transcriptome and cell proliferation of the CTX0E16 neuronal progenitor cell line. To this end I used a combination of siRNA, genome-wide expression profiling and assays of cell proliferation associated with *VRK2* knockdown.

Proliferation assays showed that reduced expression of *VRK2* decreased cell proliferation of the CTX0E16 cell line, and this was consistent with two different siRNAs and across biological replicates. This represents, to the best of my knowledge, the first direct link between VRK2 and a neural cell phenotype with potential direct relevance to schizophrenia. VRK2 has been linked to MAPK signalling which is a known modulator of cell proliferation (Zhang and Liu, 2002). VRK2 has been shown to interact directly with members of the MAPK signalling pathway such as JIP1 or KRS1 and to modulate the response mediated by this pathway (Blanco et al., 2008; Fernandez et al., 2010; Fernandez et al., 2012). Another possible explanation is that VRK2 was been shown to be able to phosphorylate molecules such as p53 and c-Jun which are known to have a role in cell cycling and proliferation (Blanco et al., 2006; Schreiber et al., 1999; Sevilla et al., 2004; Vleugel et al., 2006). Although these biochemical experiments were conducted in cell lines of cancer origin, my results suggest a similar role in brain cells. Additional studies will be needed to clarify the relevance of these known interactions in brain cells and to uncover the pathways linking VRK2 to neural progenitor cell proliferation.

The effect of VRK2 knockdown mirrors that of TCF4. TCF4 knockdown was found to change expression of genes belonging to cell signalling and proliferation pathways and I showed in Chapter 3 that knocking down TCF4 also affects cell proliferation (Chapter 3). For my study of VRK2 I performed micro-array analyses in parallel to the cell differentiation assays, providing me with the opportunity to test if the physiological effect

on cell proliferation observed after VRK2 knockdown is also supported by the microarray results.

The microarray analysis showed 186 genes to be differentially expressed in association with both of the siRNAs at $P < 0.05$ and with the same direction of change. Only ~15 would be expected by chance alone. However, I failed to find significant evidence for any gene expression change at FDR < 0.05 suggesting that these changes are small in magnitude and that this study lacks statistical power.

Interestingly, among the most genes with P < 0.05 and with same direction of change in both siRNA were CACNB2 and C10orf32, two genes at loci that have been implicated in schizophrenia by GWAS (Ripke et al., 2013).

Genes belonging to the GO terms 'intracellular', 'microsome' and 'vesicular fraction' were enriched (Bonferroni corrected $P < 0.05$) among those showing altered expression. These terms are relatively general GO terms referring to genes encoding proteins found associated with cellular and vesicular membranes. It is not straightforward to link these cellular processes and compartments with VRK2 function. However, a possible explanation is that this result is driven by the knockdown of the VRK2A isoform, which is anchored to the endoplasmic reticulum membranes by a C-terminal hydrophobic region (Blanco et al., 2006). There is evidence that VRK2A interact with scaffold proteins, which play an important role in intracellular signalling compartmentalization (Good et al., 2011). VRK2 can form complexes with scaffold and other proteins that modulate the response to stimulus mediated by MAPK signalling (Blanco et al., 2007; Blanco et al., 2008; Fernandez et al., 2012).

In addition, I looked if any of the GO categories with a nominal significance ($P < 0.05$) are associated with biological processes known to regulate cell proliferation. Four such categories, 'MAPKKK cascade' ($P = 0.0217$, Bonferroni = 1), 'JNK cascade' ($P = 0.0129$, Bonferroni = 1), 'stress-activated protein kinase signalling pathway' ($P = 0.0139$,

Bonferroni = 1), and 'cellular response to stress' (*P* = 0.0305, Bonferroni = 1) were overrepresented among the genes with altered expression after VRK2 knockdown, although these categories did not survive correction for multiple testing. Several studies have described the direct interaction between VRK2 and scaffold proteins such as JIP1 and KRS1. These together with other proteins form complexes that modulate the magnitude of the signal transmission by MAPK signalling pathways in response to hypoxia, cytokines and growth factors (Blanco et al., 2007; Blanco et al., 2008; Fernandez et al., 2010; Fernandez et al., 2012). Additionally, intracellular levels of VRK2 conditioned the direction towards a signalling pathway and the magnitude of the response mediated for it (Blanco et al., 2007). MAPK pathways play a pivotal role in signal transduction in neurons and alteration of these signalling pathways are associated with the pathology of schizophrenia (Funk et al., 2012; Kyosseva et al., 1999). Therefore, the data presented here although tentative might reflect signal from a potential mechanism for VRK2's role in schizophrenia.

The present work is the first study to show a direct effect of VRK2 on proliferation in neural cells. Interestingly, several studies have associated cell proliferation in the pathogenesis of schizophrenia (Fan et al., 2012; Reif et al., 2006; Wang et al., 2010) and there are some drugs used to treat this disease that have an effect in cell proliferation, such as valproic acid, olanzapine and others (Dodurga et al., 2014; Kimoto et al., 2011; Kurihara et al., 2014; Niu et al., 2010; Schwarz et al., 2008). This opens the possibility that VRK2 and its associated signaling pathways may mediate or be interesting targets for drugs to treat schizophrenia. Future studies could explore this possibility and help understand the mechanism on effect on cell proliferation and how it can be used to develop new therapies.

The present study provides novel data showing a functional effect of VRK2 knockdown on the transcriptional programme associated with this kinase, providing a first insight into VRK2's regulated genes and its biological function in the human brain. Contrasting with the clear effect seen on cell proliferation and the findings with TCF4, this effect is not

reflected in pathway analysis where there were no enriched GO terms after Bonferroni correction with direct links to cell proliferation. I hypothesise that this may be due to the fact that VRK2 is a kinase and its effect on transcription would therefore be indirect, and mediated by down-stream transcription factors. Indeed, VRK2 can phosphorylate transcription factor like p53 (Blanco et al., 2006) and therefore its knock down is expected to have effects on gene-expression mediated by this and other transcription factors. The effect on different transcription factors could cancel each other out and provide more room for regulatory feedback loops that could correct the deregulation induced by lower levels of VRK2. In addition, because a mixture of multiple downstream transcription factors can change expression of different signalling pathways, some of which may overlap, the genes showing changed expression may not fit as clearly on a few GO terms. More importantly, in spite of the observed effect on the transcriptome, VRK2 knockdown may have more profound effects independent of transcription, for example on phosphorylation, activation and modulation of signalling pathways without need of *de novo* transcription.

It is important to note that the analyses conducted in this study were performed at a single time point, thus providing only a snapshot of the transcriptional consequences of VRK2 knockdown in the CTX0E16 cells. It is therefore likely that some changes in gene expression arising from VRK2 knockdown went undetected.

It may be possible to increase the power to detect the effect of VRK2 knockdown on the cell transcriptome by inducing lower levels of VRK2 (more knockdown) or to perform more biological replicates. However, the clear effect on cell proliferation suggest that even small changes in transcription are enough to affect cell physiology or that other mechanisms are mediating VRK2's effect, e.g., regulation of signalling pathways by phosphorilation of its targets. Therefore, I suggest that future analyses should ideally combine cell phenotype assays (e.g., proliferation) with direct transcriptome as well as measurements of VRK2 kinase function, for instance using phosphoproteomics. This may provide additional insights and complement those findings described in this study.

Chapter 6 General Discussion

The main aim of this thesis was to characterise the molecular and cellular mechanisms by which genetic variation in *TCF4* and *VRK2* confers susceptibility to schizophrenia (Schizophrenia Psychiatric Genome-Wide Association Study, 2011; Stefansson et al., 2009). At the outset of this thesis there was very little known about the biological functions of these genes in the human brain. The findings presented in this thesis provide convincing evidence that TCF4 and VRK2 are involved in the regulation of cell proliferation of neuronal progenitor cells *in vitro*, indicating this as a possible mechanism underlying how these genes influence the pathology of schizophrenia. In addition, I found evidence for a *cis*-regulatory effect of a schizophrenia risk variant on VRK2 expression, operating during foetal brain development.

As a brief summary, the specific aims of this thesis were:

A. *In silico* characterisation of TCF4 through extensive bioinformatics analyses and data mining.
B. Test whether *VRK2* expression in the neonatal and adult human brain is influenced by the schizophrenia risk allele (rs2312147 allele G, (Steinberg et al., 2011)) and other common *cis*-regulatory variation.
C. Test whether manipulation of the endogenous expression of *TCF4* or *VRK2* has an effect on cell proliferation and to assess the effect on cell transcriptome of *VRK2* knockdown.

The bioinformatics analyses and literature review presented in Chapter 2 compiled information pointing to several potentially interesting roles of TCF4 in brain. In particular I found converging evidence of its relationship with other transcription factors known to regulate cell fate decision and neuronal differentiation. This information was important to interpret the effect on cell proliferation observed upon knockdown of *TCF4* mRNA presented in Chapter 3. These experiments were prompted by interesting findings from a

micro-array study of Dr. Bray's group where *TCF4* knockdown affected the expression of genes involved in cell cycling. However, this micro-array data lacked any functional evidence to prove that the expression differences really translated into cell physiology changes. My data show that knockdown of endogenous *TCF4* altered the proliferation of two human neural progenitor cell lines, CTX0E16 and CTX0E03. *TCF4* knockdown increased proliferation of CTX0E16 and decreased proliferation of CTX0E03. This is consistent with gene expression data obtained by Dr Nick Bray's group where genes involved in cell cycling changed expression in both cell lines but were positively correlated with *TCF4* expression in CTX0E16 cells and negatively correlated in CTX0E03 cells. The data I presented here provide evidence for a functional role of TCF4 in the regulation of cell proliferation and, interestingly, suggests that this effect may be heterogeneous across cell types.

Regarding VRK2, the data presented in chapter 4 shows the presence of active *cis*-regulatory variation affecting *VRK2* expression in dorsolateral prefrontal cortex (DLPFC) in adult human brain and two stages of foetal brain development. More interestingly, my data show an effect of genotype at rs2312147 on *VRK2* allelic expression in foetal samples from first and second trimester of gestation but not in adult DLPFC. This is the first study to provide evidence that *VRK2* is subject to variable *cis* influences on its expression. Similarly to *TCF4*, I found that knockdown of *VRK2* has an effect on cell proliferation (presented in Chapter 5). This provides the first direct link between VRK2 and a neural cell phenotype with potential relevance to schizophrenia reported to date. The micro-array profiling performed in parallel to the proliferation assay showed an effect of *VRK2* knockdown on the transcriptional programme of the cell, providing firsts insights into VRK2's downstream genes.

The effect of *TCF4* and *VRK2* knockdown on the proliferation of neural progenitor cells together with evidence for an effect of schizophrenia risk variant rs2312147 on *VRK2* expression in foetal brain provides support for the hypothesis that schizophrenia is, at least partially, a neurodevelopmental disorder, i.e., a molecular risk mechanism for

schizophrenia operating during brain development and before disease onset. (Harrison, 1997; Lewis and Levitt, 2002; Murray and Lewis, 1987; Weinberger, 1987). Interestingly, *TCF4* and *VRK2* are both highly expressed during foetal development, suggesting a role for these genes during early brain development (Figure 5.2, Chapter 5, section 5.1). Interestingly, recent integrative studies on two different but genetically related psychiatric disorders, schizophrenia and autism, have identified gene-networks harbouring genes associated with each disorder and have shown that these networks are more active during foetal brain development (Tebbenkamp et al., 2014). Thus, there is increasing evidence that schizophrenia risk variants may have a functional effect during brain development, supporting the neurodevelopmental hypothesis of the disorder. It is important to note that the different hypothesis, i.e., neurodevelopmental, neuropathological and neurochemical, for the aetiology of schizophrenia are not mutually exclusive. Identifying a developmental period where these genetic variants may be more relevant is important because it could enable more accurate models of the disorder as well as a better understanding of pathogenesis. However, we will ultimately need to understand the consequences of these early effects together with the role of the genetic variants in adult brain function to develop new effective therapies for adults with schizophrenia. Understanding the link between TCF4 and VRK2 and schizophrenia in the context of the other aetiological hypotheses remains an open challenge.

It has been previously shown that genes associated with schizophrenia are enriched within particular pathways (for example see (O'Dushlaine et al., 2011)). Based on this I hypothesize that the mechanism mediating the effect on cell proliferation of TCF4 and VRK2 is through a common pathway. *TCF4* is a downstream target of the WNT/TCF pathway and its expression has been shown to be regulated by β-catenin via its TCF binding site in cancer cell lines (Kolligs et al., 2002). Meanwhile, VRK2 has been shown to interact directly with members of the MAPK signalling pathway such as JIP1 or KRS1 and to modulate the response mediated by this pathway (Blanco et al., 2007; Blanco et al., 2008; Fernandez et al., 2010; Fernandez et al., 2012). Both pathways WNT/TCF and MAPK are important modulators of cell proliferation but I found that is not possible to

provide further evidence for a common pathway. However, there is evidence showing these two pathways can mutually regulate each other and act synergistically to regulate cellular processes in a cell dependent manner (Bikkavilli and Malbon, 2009; Cervenka et al., 2011; Guardavaccaro and Clevers, 2012; Jin et al., 2011). Therefore, at the moment I think it can only be suggested that TCF4 and VRK2 influence cell proliferation via different but functionally related pathways. However, it is noteworthy that *DISC1*, a gene that is disrupted by a chromosomal translocation co-segregating with schizophrenia and other major psychiatric disorders in a large Scottish family and one of the most widely studied genes in the context of schizophrenia, has been shown to be involved in the regulation of neuronal cell proliferation by participating in the GSK3b/β-catenin pathway (Mao et al., 2009). Therefore, TCF4 may be downstream of DISC1 and might perhaps influence proliferation through the same pathway. It is conceivable, therefore, that genes mediating the effect of schizophrenia susceptibility variants affect cell proliferation through more than a single pathway. Future studies should aim to provide more evidence for the involvement of both TCF4 and VRK2 on the WNT/TCF and MAPK pathway, attempt to clarify their interplay and how they relate to other known risk mediator genes, such as DISC1.

Several studies have suggested cell proliferation is involved in the pathology of schizophrenia using neuronal nasal epithelium and histology of post-mortem tissues (Fan et al., 2012; Reif et al., 2006). There is evidence that drugs used to treat schizophrenia and bipolar disorder (another psychotic mental illness), such as valproic acid, olanzapine and others, affect cell proliferation (Dodurga et al., 2014; Kimoto et al., 2011; Kurihara et al., 2014; Niu et al., 2010; Schwarz et al., 2008). Interestingly, these drugs seem to target the same proliferation pathways linked to TCF4 and VRK2. For example, valproic acid can trigger phosphorylation of ERK interfering with the MAPK signalling pathway (Hao et al., 2004; Jung et al., 2008; Yuan et al., 2001). Additionally, valproic acid has been show to affect the WNT/ β-catenin signalling pathway by inhibiting GSK3β (Chen et al., 1999; Kao et al., 2013; Sintoni et al., 2013). Other antipsychotics have been reported to be involved in the regulation of the WNT pathway (Sutton and Rushlow, 2011). This opens

the possibility that VRK2 or TCF4 and its associated signaling pathways may mediate or be interesting targets for drugs to treat schizophrenia. I think it might the possible to combine the cellular models I have worked on to study TCF4 and VRK2 with animal models to devise a framework on which to test potential new drugs.

My argument is as follows; some drugs use to treat schizophrenia have been shown to have an effect in proliferation and VRK2 and TCF4 have an effect on proliferation, as shown in this thesis. Moreover, the evidence also suggests an overlap on the pathways where these genes and some drugs act. Thus, it can be hypothesized that the risk and therapeutic effect of these genes and drugs, respectively, converge on the same pathways. It is important to note that the drugs are used in adulthood implying, at least, two things. First, the effect on proliferation mediated by TCF4 and VRK2 might be present in adulthood in the few brain regions where neurogenesis is still active, such as the subventricular zone (SVZ) and dentate gyrus (Reif et al., 2007). A second possibility is that the disease risk mediated by these genes is not increasing risk in adulthood but the drugs can compensate some of the pathology produced during foetal development and or pathology that is a consequence of the disease itself. Regardless of which model is correct, the fact that drugs do indeed improve schizophrenia symptoms shows that drugs targeting cell proliferation can be effective in treating the disease. This is in spite of us not fully understanding all the disease mechanisms nor the mechanisms of drug action or whether schizophrenia genetic susceptibility is mostly mediated during foetal development. A interesting example of this have been shown in a KO mouse model of Fragile-X syndrome, where drugs can revert the effect of the mutation in adult mice in spite being a *bona fide* example of a neurodevelopmental disorder (Gandhi et al., 2014; Sun et al., 2014).

I translate this argument on the following idea. We can use the effect on cell proliferation on cell lines as a functional signature useful to identify new drugs to treat schizophrenia. Given the evidence presented above I suggest the knockdowns of *TCF4* and *VRK2* could be useful to this end. Therefore, we could test drugs on these cellular models and look for a reversal of the gene knockdowns on cell proliferation. If this is seen, we could think that

those drugs could have a therapeutic effect on schizophrenia. For example, known drugs targeting the WNT/TCF and MAPK pathway could be tested together for reversal of the effect on cell proliferation seen upon TCF4 or VRK2 knockdown. Targeting the same pathways where TCF4 and VRK2 act is not a requirement but could prioritize for more interesting drugs. Brzozka et al., (Brzózka et al., 2010) described a transgenic mouse for TCF4 that presents cognitive and sensorimotor impairments. This animal model could be used to test the drugs affecting proliferation in the cellular models to see if they bring about an improvement in behavioral outcomes. In addition, VRK2 is a kinase with reduced kinase promiscuity (Fedorov et al., 2007) making it the ideal candidate for development of specific inhibitors and for which selective drugs could be tested, perhaps also based on their effect on cell proliferation in the neural cell line model.

I acknowledge limitations of the experimental models and results I have presented but I also think this thesis highlights additional potentially interesting avenues for future work as well as some important considerations. Data presented in Chapter 4 show the presence of *cis*-acting influences affecting *VRK2* expression in dorsolateral prefrontal cortex (DLPFC) in adult human brain. However, the genotype at rs2312147, the SNP associated with schizophrenia, did not have a major effect on *VRK2* allelic expression in this region of the adult brain. It is important to consider that effects of *cis*-regulatory variation can differ across brain regions (Buonocore et al., 2010). Given the evidence, presented in Chapter 5, that *VRK2* knockdown affects neural cell proliferation, it may be informative to test effects of the GWAS risk variant on *VRK2* expression in an adult brain region where neurogenesis continues, such as the dentate gyrus of the hippocampus. It is also important to point out that assessment of allelic expression carried out in the present study was limited to a single gene – VRK2. However further studies using RNA-seq technology will allow assessment of allelic expression on a genome-wide scale (Pastinen, 2010).

In the present study I provided evidence for the biological effect of *TCF4* and *VRK2* knockdown in neural progenitor cell lines derived from human foetal brain. It is important

to consider some limitations of this model; for example, while celular models are a good approach to test the effect of one gene, they fail to capture all the complexity of the brain and psychiatric disorders. Furthermore, it is important to point out that effects observed in this study would be specific to cycling cells like neural progenitors. Because they are neural progenitor cells derived from human foetal brain, the cells used in these studies might not be relevant to mature cells of the adult brain.

Additionally, my results also showed that the biological effect of TCF4 and VRK2 is heterogeneous in respect to cell lines and genes. Thus, it seems that in order to gain a more detailed understanding we may need to explore the use of additional cell lines. For instance, this could be achieved with studies in mature neuronal cells differentiated from these progenitors. One could also think of more complex but also realistic experimental models. For example those in which different mature cell lineages generate active functional synapses on which electrochemical properties of synapses and other cell physiological phenotypes could be measured. Perhaps in the future it will be possible to induce knockdown in cerebral organoids. Cerebral organoid permits *in vitro* assembly of a human embryonic brain rudiment that recapitulates the developing human cerebrum and yield mature cortical neuron subtypes (Lancaster et al., 2013; Muzio and Consalez, 2013). Regardless, of the model I think it might be necessary to use additional and differentiated cell lineages and if possible study the effect of knockdown (or other intervention) on their functional interactions.

I believe functional experiments aiming to identify the biological processes likely affected by alleles confering risk to schizophrenia may also have implication for the understanding of future genetic studies both in populations and in the case of personalised medicine. One can think of using the knowledge of the biological processes to prioritize genetic variants implicated in a disorder. For example, to date analyses of rare variants have been restricted to burden tests grouping all variants together. In order to link susceptibility alleles with biological mechanism or to practice stratified medicine we need to identify individual risk variants. Recognizing true susceptibility rare variation among the sea of

rare variants in the human genome is currently a major challenge (MacArthur et al., 2014). An interesting possibility is that the effect of both rare and common variants will converge on common biological processes. Therefore, as we characterise the biological role of genes in additional risk loci this information will serve as a framework to sieve human rare genetic variation and retain the most likely rare susceptibility alleles for further studies. For example, in a patient with schizophrenia and whole genome sequencing we could pay more attention or give more weight to rare potentially deleterious variation in genes known to affect cell proliferation or to be highly expressed during mid-foetal brain development. These ideas are already being applied to other human disease to prioritise variants for further research, for example to autism (Liu et al., 2014), and might be important for the interpretation of personal genomes and quantitative personalised medicine.

At the moment, the potential links between TCF4 and VRK with the WNT/TCF and MAPK pathways, respectively, are based on biochemical evidence gathered from other human cell types and other species. I have used this evidence to speculate on the possible mechanism underlying the effects on cell proliferation. However, it remains for future work to validate and extend knowledge of these biochemical interactions. This may be especially interesting as one attempts to validate links with other genes involved in schizophrenia that also affect cell proliferation, like DISC1.

Finally, I used transcriptome profiling to better understand the molecular effect of VRK2 knockdown. I think this is a valuable approach since it is genome-wide, requires no a priori hypotheses and enables comparison with other data generated at Dr. Bray's lab on the same cell lines and other genes. However, given that VRK2 is a kinase it would be interesting to use other experimental approaches like phosphoproteomics to better understand VRK2's kinase function and how it relates to MAPK and WNT/TCF pathways and genes in other schizophrenia associated loci. These studies combined with cell phenotype assays (e.g., proliferation) and direct trasncriptome profiling will complement

my findings and provide additional insights into VRK2's regulated genes and targets and its biological function in the human brain.

In summary, the series of studies presented in this thesis provide experimental evidence for the biological effect of TCF4 and VRK2 on cell proliferation in human neural progenitor cells, and evidence that schizophrenia risk variation at the *VRK2* locus has a *cis*-effect on *VRK2* expression during foetal brain development. These findings can be interpreted in the context of and support the neurodevelopmental hypothesis of schizophrenia. I highlight interesting implications of my results, some of which warrant future research to help dissect the molecular aetiology of schizophrenia and to develop models to bring new treatments to patients with this devastating disease.

REFERENCES

Aberg, K.A., Liu, Y., Bukszar, J., McClay, J.L., Khachane, A.N., Andreassen, O.A., Blackwood, D., Corvin, A., Djurovic, S., Gurling, H.*, et al.* (2013). A comprehensive family-based replication study of schizophrenia genes. JAMA psychiatry *70*, 573-581.

Abu-Elneel, K., Liu, T., Gazzaniga, F.S., Nishimura, Y., Wall, D.P., Geschwind, D.H., Lao, K., and Kosik, K.S. (2008). Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics *9*, 153-161.

Abu-Khalil, A., Fu, L., Grove, E.A., Zecevic, N., and Geschwind, D.H. (2004). Wnt genes define distinct boundaries in the developing human brain: implications for human forebrain patterning. J Comp Neurol *474*, 276-288.

Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S., and Kageyama, R. (1995). A mammalian helix-loop-helix factor structurally related to the product of Drosophila proneural gene atonal is a positive transcriptional regulator expressed in the developing nervous system. J Biol Chem *270*, 8730-8738.

Akbarian, S., Kim, J.J., Potkin, S.G., Hagman, J.O., Tafazzoli, A., Bunney, W.E., Jr., and Jones, E.G. (1995). Gene expression for glutamic acid decarboxylase is reduced without loss of neurons in prefrontal cortex of schizophrenics. Arch Gen Psychiatry *52*, 258-266.

Albanna, A., Choudhry, Z., Harvey, P.O., Fathalli, F., Cassidy, C., Sengupta, S.M., Iyer, S.N., Rho, A., Lepage, M., Malla, A.*, et al.* (2014). TCF4 gene polymorphism and cognitive performance in patients with first episode psychosis. Schizophr Res *152*, 124-129.

Alkelai, A., Greenbaum, L., Lupoli, S., Kohn, Y., Sarner-Kanyas, K., Ben-Asher, E., Lancet, D., Macciardi, F., and Lerer, B. (2012). Association of the type 2 diabetes mellitus susceptibility gene, TCF7L2, with schizophrenia in an Arab-Israeli family sample. PLoS One *7*, e29228.

Almlof, J.C., Lundmark, P., Lundmark, A., Ge, B., Maouche, S., Goring, H.H., Liljedahl, U., Enstrom, C., Brocheton, J., Proust, C.*, et al.* (2012). Powerful identification of cis-

regulatory SNPs in human primary monocytes using allele-specific gene expression. PLoS One *7*, e52260.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Altshuler, D., Daly, M.J., and Lander, E.S. (2008). Genetic mapping in human disease. Science *322*, 881-888.

Amaral, F.C., Torres, N., Saggioro, F., Neder, L., Machado, H.R., Silva, W.A., Jr., Moreira, A.C., and Castro, M. (2009). MicroRNAs differentially expressed in ACTH-secreting pituitary tumors. J Clin Endocrinol Metab *94*, 320-323.

Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., Plouin, P., Carter, N.P., Lyonnet, S., Munnich, A.*, et al.* (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet *80*, 988-993.

Andreasen, N.C. (1995). Symptoms, signs, and diagnosis of schizophrenia. Lancet *346*, 477-481.

Appaiah, H., Bhat-Nakshatri, P., Mehta, R., Thorat, M., Badve, S., and Nakshatri, H. (2010). ITF2 is a target of CXCR4 in MDA-MB-231 breast cancer cells and is associated with reduced survival in estrogen receptor-negative breast cancer. Cancer biology & therapy *10*, 600-614.

Association, A.P. (2000). Diagnostic and statistical manual of mental disorders. Fourth Edition. Washington, DC.

Badner, J.A., and Gershon, E.S. (2002). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. Mol Psychiatry *7*, 405-411.

Baratz, K.H., Tosakulwong, N., Ryu, E., Brown, W.L., Branham, K., Chen, W., Tran, K.D., Schmid-Kubista, K.E., Heckenlively, J.R., Swaroop, A.*, et al.* (2010). E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med *363*, 1016-1024.

Barnes, M.R., Huxley-Jones, J., Maycox, P.R., Lennon, M., Thornber, A., Kelly, F., Bates, S., Taylor, A., Reid, J., Jones, N.*, et al.* (2011). Transcription and pathway analysis of the

superior temporal cortex and anterior prefrontal cortex in schizophrenia. J Neurosci Res *89*, 1218-1227.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.

Beasley, C.L., and Reynolds, G.P. (1997). Parvalbumin-immunoreactive neurons are reduced in the prefrontal cortex of schizophrenics. Schizophr Res *24*, 349-355.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y., and Pritchard, J.K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol *12*, R10.

Ben-Arie, N., Bellen, H.J., Armstrong, D.L., McCall, A.E., Gordadze, P.R., Guo, Q., Matzuk, M.M., and Zoghbi, H.Y. (1997). Math1 is essential for genesis of cerebellar granule neurons. Nature *390*, 169-172.

Ben-Arie, N., McCall, A.E., Berkman, S., Eichele, G., Bellen, H.J., and Zoghbi, H.Y. (1996). Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis. Hum Mol Genet *5*, 1207-1216.

Benes, F.M. (1993). Neurobiological investigations in cingulate cortex of schizophrenic brain. Schizophr Bull *19*, 537-549.

Beneyto, M., Kristiansen, L.V., Oni-Orisan, A., McCullumsmith, R.E., and Meador-Woodruff, J.H. (2007). Abnormal glutamate receptor expression in the medial temporal lobe in schizophrenia and mood disorders. Neuropsychopharmacology *32*, 1888-1902.

Bergen, S.E., O'Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., Moran, J.L., Chambert, K.D., Handsaker, R.E., Backlund, L.*, et al.* (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry *17*, 880-886.

Bergqvist, I., Eriksson, M., Saarikettu, J., Eriksson, B., Corneliussen, B., Grundström, T., and Holmberg, D. (2000). The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. European Journal of Immunology *30*, 2857-2863.

Berman, K.F., Illowsky, B.P., and Weinberger, D.R. (1988). Physiological dysfunction of dorsolateral prefrontal cortex in schizophrenia. IV. Further evidence for regional and behavioral specificity. Arch Gen Psychiatry *45*, 616-622.

Betcheva, E.T., Yosifova, A.G., Mushiroda, T., Kubo, M., Takahashi, A., Karachanak, S.K., Zaharieva, I.T., Hadjidekova, S.P., Dimova, II, Vazharova, R.V.*, et al.* (2013). Whole-genome-wide association study in the Bulgarian population reveals HHAT as schizophrenia susceptibility gene. Psychiatr Genet *23*, 11-19.

Bhattacharya, A., and Baker, N.E. (2011). A network of broadly expressed HLH genes regulates tissue-specific cell fates. Cell *147*, 881-892.

Bian, F., Liu, W., Yoon, K.C., Lu, R., Zhou, N., Ma, P., Pflugfelder, S.C., and Li, D.Q. (2010). Molecular signatures and biological pathway profiles of human corneal epithelial progenitor cells. Int J Biochem Cell Biol *42*, 1142-1153.

Biegert, A., Mayer, C., Remmert, M., Soding, J., and Lupas, A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. Nucleic Acids Res *34*, W335-339.

Bikkavilli, R.K., and Malbon, C.C. (2009). Mitogen-activated protein kinases and Wnt/beta-catenin signaling: Molecular conversations among signaling pathways. Communicative & integrative biology *2*, 46-49.

Black, J.E., Kodish, I.M., Grossman, A.W., Klintsova, A.Y., Orlovskaya, D., Vostrikov, V., Uranova, N., and Greenough, W.T. (2004). Pathology of layer V pyramidal neurons in the prefrontal cortex of patients with schizophrenia. Am J Psychiatry *161*, 742-744.

Blake, D.J., Forrest, M., Chapman, R.M., Tinsley, C.L., O'Donovan, M.C., and Owen, M.J. (2010). TCF4, schizophrenia, and Pitt-Hopkins Syndrome. Schizophr Bull *36*, 443-447.

Blanco, S., Klimcakova, L., Vega, F.M., and Lazo, P.A. (2006). The subcellular localization of vaccinia-related kinase-2 (VRK2) isoforms determines their different effect on p53 stability in tumour cell lines. FEBS J *273*, 2487-2504.

Blanco, S., Santos, C., and Lazo, P.A. (2007). Vaccinia-related kinase 2 modulates the stress response to hypoxia mediated by TAK1. Mol Cell Biol *27*, 7273-7283.

Blanco, S., Sanz-Garcia, M., Santos, C.R., and Lazo, P.A. (2008). Modulation of interleukin-1 transcriptional response by the interaction between VRK2 and the JIP1 scaffold protein. PLoS ONE *3*, e1660.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol *Chapter 19*, Unit 19 10 11-21.

Boyer, P., Phillips, J.L., Rousseau, F.L., and Ilivitsky, S. (2007). Hippocampal abnormalities and memory deficits: new evidence of a strong pathophysiological link in schizophrenia. Brain Res Rev *54*, 92-112.

Brandon, N.J., and Sawa, A. (2011). Linking neurodevelopmental and synaptic theories of mental illness through DISC1. Nature reviews Neuroscience *12*, 707-722.

Bray, N.J., Buckland, P.R., Owen, M.J., and O'Donovan, M.C. (2003). Cis-acting variation in the expression of a high proportion of genes in human brain. Hum Genet *113*, 149-153.

Bray, N.J., Kapur, S., and Price, J. (2012). Investigating schizophrenia in a "dish": possibilities, potential and limitations. World Psychiatry *11*, 153-155.

Bray, N.J., and O'Donovan, M.C. (2006). Investigating cis-acting regulatory variation using assays of relative allelic expression. Psychiatr Genet *16*, 173-177.

Breier, A., Su, T.P., Saunders, R., Carson, R.E., Kolachana, B.S., de Bartolomeis, A., Weinberger, D.R., Weisenfeld, N., Malhotra, A.K., Eckelman, W.C.*, et al.* (1997). Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. Proc Natl Acad Sci U S A *94*, 2569-2574.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D.*, et al.* (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221-225.

Brockschmidt, A., Filippi, A., Charbel Issa, P., Nelles, M., Urbach, H., Eter, N., Driever, W., and Weber, R.G. (2011). Neurologic and ocular phenotype in Pitt-Hopkins syndrome and a zebrafish model. Hum Genet *130*, 645-655.

Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., Frenck, W., Radlwimmer, B., Lichter, P., Engels, H.*, et al.* (2007). Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. Hum Mol Genet *16*, 1488-1494.

Brody, J.R., and Witkiewicz, A.K. (2010). CXCR4 signaling identifies a role for IFT2 in ER-negative breast cancers. Cancer biology & therapy *10*, 615-616.

Brown, A.S., and Patterson, P.H. (2011). Maternal infection and schizophrenia: implications for prevention. Schizophr Bull *37*, 284-290.

Brown, A.S., Susser, E.S., Butler, P.D., Richardson Andrews, R., Kaufmann, C.A., and Gorman, J.M. (1996). Neurobiological plausibility of prenatal nutritional deprivation as a risk factor for schizophrenia. The Journal of nervous and mental disease *184*, 71-85.

Brzózka, M.M., Radyushkin, K., Wichert, S.P., Ehrenreich, H., and Rossner, M.J. (2010). Cognitive and Sensorimotor Gating Impairments in Transgenic Mice Overexpressing the Schizophrenia Susceptibility Gene Tcf4 in the Brain. Biological Psychiatry *68*, 33-40.

Buckley, P.F., Miller, B.J., Lehrer, D.S., and Castle, D.J. (2009). Psychiatric comorbidities and schizophrenia. Schizophr Bull *35*, 383-402.

Buonocore, F., Hill, M.J., Campbell, C.D., Oladimeji, P.B., Jeffries, A.R., Troakes, C., Hortobagyi, T., Williams, B.P., Cooper, J.D., and Bray, N.J. (2010). Effects of cis-regulatory variation differ across regions of the adult human brain. Hum Mol Genet *19*, 4490-4496.

Byne, W., Kidkardnee, S., Tatusov, A., Yiannoulos, G., Buchsbaum, M.S., and Haroutunian, V. (2006). Schizophrenia-associated reduction of neuronal and oligodendrocyte numbers in the anterior principal thalamic nucleus. Schizophr Res *85*, 245-253.

Cannon, M., Jones, P.B., and Murray, R.M. (2002). Obstetric complications and schizophrenia: historical and meta-analytic review. Am J Psychiatry *159*, 1080-1092.

Cardno, A.G., Rijsdijk, F.V., Sham, P.C., Murray, R.M., and McGuffin, P. (2002). A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry *159*, 539-545.

Care, A., Catalucci, D., Felicetti, F., Bonci, D., Addario, A., Gallo, P., Bang, M.L., Segnalini, P., Gu, Y., Dalton, N.D.*, et al.* (2007). MicroRNA-133 controls cardiac hypertrophy. Nat Med *13*, 613-618.

Carlsson, A., Waters, N., Holm-Waters, S., Tedroff, J., Nilsson, M., and Carlsson, M.L. (2001). Interactions between monoamines, glutamate, and GABA in schizophrenia: new evidence. Annual review of pharmacology and toxicology *41*, 237-260.

Casanova, M.F., Kreczmanski, P., Trippe, J., 2nd, Switala, A., Heinsen, H., Steinbusch, H.W., and Schmitz, C. (2008). Neuronal distribution in the neocortex of schizophrenic patients. Psychiatry Res *158*, 267-277.

Cervantes-Barragan, L., Lewis, K.L., Firner, S., Thiel, V., Hugues, S., Reith, W., Ludewig, B., and Reizis, B. (2012). Plasmacytoid dendritic cells control T-cell response to chronic viral infection. Proc Natl Acad Sci U S A *109*, 3012-3017.

Cervenka, I., Wolf, J., Masek, J., Krejci, P., Wilcox, W.R., Kozubik, A., Schulte, G., Gutkind, J.S., and Bryja, V. (2011). Mitogen-activated protein kinases promote WNT/beta-catenin signaling via phosphorylation of LRP6. Mol Cell Biol *31*, 179-189.

Chabchoub, E., Vermeesch, J.R., de Ravel, T., de Cock, P., and Fryns, J.P. (2008). The facial dysmorphy in the newly recognised microdeletion 2p15-p16.1 refined to a 570 kb region in 2p15. J Med Genet *45*, 189-192.

Chandrasekar, V., and Dreyer, J.L. (2009). microRNAs miR-124, let-7d and miR-181a regulate cocaine-induced plasticity. Mol Cell Neurosci *42*, 350-362.

Chen, G., Huang, L.D., Jiang, Y.M., and Manji, H.K. (1999). The mood-stabilizing agent valproate inhibits the activity of glycogen synthase kinase-3. J Neurochem *72*, 1327-1330.

Cheung, V.G., and Spielman, R.S. (2009). Genetics of human gene expression: mapping DNA variants that influence gene expression. Nature reviews Genetics *10*, 595-604.

Chevallier, N.L., Soriano, S., Kang, D.E., Masliah, E., Hu, G., and Koo, E.H. (2005). Perturbed neurogenesis in the adult hippocampus associated with presenilin-1 A246E mutation. The American journal of pathology *167*, 151-159.

Chiaramello, A., Soosaar, A., Neuman, T., and Zuber, M.X. (1995). Differential expression and distinct DNA-binding specificity of ME1a and ME2 suggest a unique role during differentiation and neuronal plasticity. Brain research Molecular brain research *29*, 107-118.

Chua, S.E., and Murray, R.M. (1996). The neurodevelopmental theory of schizophrenia: evidence concerning structure and neuropsychology. Annals of medicine *28*, 547-555.

Cisse, B., Caton, M.L., Lehner, M., Maeda, T., Scheu, S., Locksley, R., Holmberg, D., Zweier, C., den Hollander, N.S., Kant, S.G.*, et al.* (2008). Transcription factor E2-2 is an

essential and specific regulator of plasmacytoid dendritic cell development. Cell *135*, 37-48.

Clarke, M.C., Harley, M., and Cannon, M. (2006). The role of obstetric events in schizophrenia. Schizophr Bull *32*, 3-8.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B.*, et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R.*, et al.* (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature *478*, 519-523.

Collins, F.S. (1995). Positional cloning moves from perditional to traditional. Nat Genet *9*, 347-350.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P.*, et al.* (2010). Origins and functional impact of copy number variation in the human genome. Nature *464*, 704-712.

Consortium, E., Consortium, E.M., Steffens, M., Leu, C., Ruppert, A.K., Zara, F., Striano, P., Robbiano, A., Capovilla, G., Tinuper, P.*, et al.* (2012). Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet *21*, 5359-5372.

Conti, A., Aguennouz, M., La Torre, D., Tomasello, C., Cardali, S., Angileri, F.F., Maio, F., Cama, A., Germano, A., Vita, G.*, et al.* (2009). miR-21 and 221 upregulation and miR-181b downregulation in human grade II-IV astrocytic tumors. J Neurooncol *93*, 325-332.

Corneliussen, B., Holm, M., Waltersson, Y., Onions, J., Hallberg, B., Thornell, A., and Grundstrom, T. (1994). Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. Nature *368*, 760-764.

Corneliussen, B., Thornell, A., Hallberg, B., and Grundstrom, T. (1991). Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. Journal of virology *65*, 6084-6093.

Corvin, A., Craddock, N., and Sullivan, P.F. (2010). Genome-wide association studies: a primer. Psychological medicine *40*, 1063-1077.

Cotter, D., Kerwin, R., al-Sarraji, S., Brion, J.P., Chadwich, A., Lovestone, S., Anderton, B., and Everall, I. (1998). Abnormalities of Wnt signalling in schizophrenia--evidence for neurodevelopmental abnormality. Neuroreport *9*, 1379-1383.

Cowles, C.R., Hirschhorn, J.N., Altshuler, D., and Lander, E.S. (2002). Detection of regulatory variation in mouse genes. Nat Genet *32*, 432-437.

Creese, I., Burt, D.R., and Snyder, S.H. (1976). Dopamine receptors and average clinical doses. Science *194*, 546.

Cross-Disorder Group of the Psychiatric Genomics, C., and Genetic Risk Outcome of Psychosis, C. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371-1379.

Crump, C., Winkleby, M.A., Sundquist, K., and Sundquist, J. (2013). Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study. Am J Psychiatry *170*, 324-333.

de Leeuw, N., Pfundt, R., Koolen, D.A., Neefs, I., Scheltinga, I., Mieloo, H., Sistermans, E.A., Nillesen, W., Smeets, D.F., de Vries, B.B.*, et al.* (2008). A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. J Med Genet *45*, 122-124.

de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A.*, et al.* (2009). Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. Hum Mutat *30*, 669-676.

de Pooter, R.F., and Kee, B.L. (2010). E proteins and the regulation of early lymphocyte development. Immunological reviews *238*, 93-109.

Dermitzakis, E.T. (2008). From gene expression to disease risk. Nat Genet *40*, 492-493.

Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Gutierrez Arcelus, M., Sekowska, M.*, et al.* (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. Science *325*, 1246-1250.

Dimas, A.S., Stranger, B.E., Beazley, C., Finn, R.D., Ingle, C.E., Forrest, M.S., Ritchie, M.E., Deloukas, P., Tavare, S., and Dermitzakis, E.T. (2008). Modifier effects between regulatory and protein-coding variation. PLoS Genet *4*, e1000244.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M.*, et al.* (2007a). A genome-wide association study of global gene expression. Nat Genet *39*, 1202-1207.

Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C.C., Taylor, J., Burnett, E., Gut, I., Farrall, M.*, et al.* (2007b). A genome-wide association study of global gene expression. Nature Genetics *39*, 1202-1207.

Dodurga, Y., Gundogdu, G., Tekin, V., Koc, T., Satiroglu-Tufan, N.L., Bagci, G., and Kucukatay, V. (2014). Valproic acid inhibits the proliferation of SHSY5Y neuroblastoma cancer cells by downregulating URG4/URGCP and CCND1 gene expression. Molecular biology reports.

Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547-1548.

Dudbridge, F., and Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. Genet Epidemiol *32*, 227-234.

Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. J Biomed Inform *44*, 839-847.

Dworkin, S., and Mantamadiotis, T. (2010). Targeting CREB signalling in neurogenesis. Expert Opin Ther Targets *14*, 869-879.

Eastwood, S.L. (2004). The synaptic pathology of schizophrenia: is aberrant neurodevelopment and plasticity to blame? Int Rev Neurobiol *59*, 47-72.

Eastwood, S.L., and Harrison, P.J. (2010). Markers of glutamate synaptic transmission and plasticity are increased in the anterior cingulate cortex in bipolar disorder. Biol Psychiatry *67*, 1010-1016.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Einarson, M.B., and Chao, M.V. (1995). Regulation of Id1 and its association with basic helix-loop-helix proteins during nerve growth factor-induced differentiation of PC12 cells. Mol Cell Biol *15*, 4175-4183.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc *2*, 953-971.

Engel, I., and Murre, C. (2001). The function of E- and Id proteins in lymphocyte development. Nat Rev Immunol *1*, 193-199.

Erlenmeyer-Kimling, L., Rock, D., Roberts, S.A., Janal, M., Kestenbaum, C., Cornblatt, B., Adamo, U.H., and Gottesman, II (2000). Attention, memory, and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High-Risk Project. Am J Psychiatry *157*, 1416-1422.

Ernst, A., Campos, B., Meier, J., Devens, F., Liesenberg, F., Wolter, M., Reifenberger, G., Herold-Mende, C., Lichter, P., and Radlwimmer, B. (2010). De-repression of CTGF via the miR-17-92 cluster upon differentiation of human glioblastoma spheroid cultures. Oncogene *29*, 3411-3422.

Esashi, E., and Watowich, S.S. (2009). Dendritic cells: transcriptional control of plasmacytoid dendritic cell development by E2-2. Immunol Cell Biol *87*, 1-2.

Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B.*, et al.* (2007). Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res *17*, 898-909.

Evangelisti, C., Florian, M.C., Massimi, I., Dominici, C., Giannini, G., Galardi, S., Bue, M.C., Massalini, S., McDowell, H.P., Messi, E.*, et al.* (2009). MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness. FASEB J *23*, 4276-4287.

Fan, Y., Abrahamsen, G., McGrath, J.J., and Mackay-Sim, A. (2012). Altered cell cycle dynamics in schizophrenia. Biol Psychiatry *71*, 129-135.

Farah, M.H., Olson, J.M., Sucic, H.B., Hume, R.I., Tapscott, S.J., and Turner, D.L. (2000). Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. Development *127*, 693-702.

Fatemi, S.H., and Folsom, T.D. (2009). The neurodevelopmental hypothesis of schizophrenia, revisited. Schizophr Bull *35*, 528-548.

Fedorov, O., Marsden, B., Pogacic, V., Rellos, P., Muller, S., Bullock, A.N., Schwaller, J., Sundstrom, M., and Knapp, S. (2007). A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. Proc Natl Acad Sci U S A *104*, 20523-20528.

Felix, T.M., Petrin, A.L., Sanseverino, M.T., and Murray, J.C. (2010). Further characterization of microdeletion syndrome involving 2p15-p16.1. Am J Med Genet A *152A*, 2604-2608.

Fernandez, I.F., Blanco, S., Lozano, J., and Lazo, P.A. (2010). VRK2 inhibits mitogen-activated protein kinase signaling and inversely correlates with ErbB2 in human breast cancer. Mol Cell Biol *30*, 4687-4697.

Fernandez, I.F., Perez-Rivas, L.G., Blanco, S., Castillo-Dominguez, A.A., Lozano, J., and Lazo, P.A. (2012). VRK2 anchors KSR1-MEK1 to endoplasmic reticulum forming a macromolecular complex that compartmentalizes MAPK signaling. Cellular and molecular life sciences : CMLS *69*, 3881-3893.

Flora, A., Garcia, J.J., Thaller, C., and Zoghbi, H.Y. (2007). The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proceedings of the National Academy of Sciences *104*, 15382-15387.

Forrest, M.P., Waite, A.J., Martin-Rendon, E., and Blake, D.J. (2013). Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. PLoS One *8*, e73169.

Frankle, W.G., and Laruelle, M. (2002). Neuroreceptor imaging in psychiatric disorders. Annals of nuclear medicine *16*, 437-446.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res *32*, W273-279.

Freyberg, Z., Ferrando, S.J., and Javitch, J.A. (2010). Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. Am J Psychiatry *167*, 388-396.

Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelmann, L.*, et al.* (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry *13*, 261-266.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M.*, et al.* (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179-184.

Fu, C.H., and McGuire, P.K. (1999). Functional neuroimaging in psychiatry. Philosophical transactions of the Royal Society of London Series B, Biological sciences *354*, 1359-1370.

Fu, H., Cai, J., Clevers, H., Fast, E., Gray, S., Greenberg, R., Jain, M.K., Ma, Q., Qiu, M., Rowitch, D.H.*, et al.* (2009). A genome-wide screen for spatially restricted expression patterns identifies transcription factors that regulate glial development. J Neurosci *29*, 11399-11408.

Funk, A.J., McCullumsmith, R.E., Haroutunian, V., and Meador-Woodruff, J.H. (2012). Abnormal activity of the MAPK- and cAMP-associated signaling pathways in frontal cortical areas in postmortem brain in schizophrenia. Neuropsychopharmacology *37*, 896-905.

Furumura, M., Potterf, S.B., Toyofuku, K., Matsunaga, J., Muller, J., and Hearing, V.J. (2001). Involvement of ITF2 in the transcriptional regulation of melanogenic genes. J Biol Chem *276*, 28147-28154.

Gandhi, R.M., Kogan, C.S., and Messier, C. (2014). 2-Methyl-6-(phenylethynyl) pyridine (MPEP) reverses maze learning and PSD-95 deficits in Fmr1 knock-out mice. Frontiers in cellular neuroscience *8*, 70.

Gao, Z., Ure, K., Ding, P., Nashaat, M., Yuan, L., Ma, J., Hammer, R.E., and Hsieh, J. (2011). The master negative regulator REST/NRSF controls adult neurogenesis by restraining the neurogenic program in quiescent stem cells. J Neurosci *31*, 9772-9786.

Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Geschwind, D.H., and Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. Nature *461*, 908-915.

Ghosh, H.S., Cisse, B., Bunin, A., Lewis, K.L., and Reizis, B. (2010). Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. Immunity *33*, 905-916.

Giegling, I., Genius, J., Benninghoff, J., and Rujescu, D. (2010). Genetic findings in schizophrenia patients related to alterations in the intracellular Ca-homeostasis. Progress in neuro-psychopharmacology & biological psychiatry *34*, 1375-1380.

Glessner, J.T., Reilly, M.P., Kim, C.E., Takahashi, N., Albano, A., Hou, C., Bradfield, J.P., Zhang, H., Sleiman, P.M., Flory, J.H.*, et al.* (2010). Strong synaptic transmission impact by copy number variations in schizophrenia. Proc Natl Acad Sci U S A *107*, 10584-10589.

Gohlke, J.M., Armant, O., Parham, F.M., Smith, M.V., Zimmer, C., Castro, D.S., Nguyen, L., Parker, J.S., Gradwohl, G., Portier, C.J.*, et al.* (2008). Characterization of the proneural gene regulatory network during mouse telencephalon development. BMC Biol *6*, 15.

Goldfarb, A.N., Lewandowska, K., and Shoham, M. (1996). Determinants of helix-loop-helix dimerization affinity. Random mutational analysis of SCL/tal. J Biol Chem *271*, 2683-2688.

Gonzalez-Perez, A., and Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet *88*, 440-449.

Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. Science *332*, 680-686.

Gresch, O., Engel, F.B., Nesic, D., Tran, T.T., England, H.M., Hickman, E.S., Korner, I., Gan, L., Chen, S., Castro-Obregon, S.*, et al.* (2004). New non-viral method for gene transfer into primary cells. Methods *33*, 151-163.

Gross, J., Grimm, O., Meyer, J., and Lesch, K.P. (2003). [The cadherin hypothesis of schizophrenia]. Fortschr Neurol Psychiatr *71*, 84-88.

Guardavaccaro, D., and Clevers, H. (2012). Wnt/beta-catenin and MAPK signaling: allies and enemies in different battlefields. Science signaling *5*, pe15.

Guidotti, A., Auta, J., Davis, J.M., Di-Giorgi-Gerevini, V., Dwivedi, Y., Grayson, D.R., Impagnatiello, F., Pandey, G., Pesold, C., Sharma, R.*, et al.* (2000). Decrease in reelin and glutamic acid decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study. Arch Gen Psychiatry *57*, 1061-1069.

Gustavsson, P., Kimber, E., Wahlstrom, J., and Anneren, G. (1999). Monosomy 18q syndrome and atypical Rett syndrome in a girl with an interstitial deletion (18)(q21.1q22.3). Am J Med Genet *82*, 348-351.

Hakak, Y., Walker, J.R., Li, C., Wong, W.H., Davis, K.L., Buxbaum, J.D., Haroutunian, V., and Fienberg, A.A. (2001). Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proc Natl Acad Sci U S A *98*, 4746-4751.

Halberstadt, A.L. (1995). The phencyclidine-glutamate model of schizophrenia. Clinical neuropharmacology *18*, 237-249.

Hansen, T., Ingason, A., Djurovic, S., Melle, I., Fenger, M., Gustafsson, O., Jakobsen, K.D., Rasmussen, H.B., Tosato, S., Rietschel, M.*, et al.* (2011). At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. Biol Psychiatry *70*, 59-63.

Hao, Y., Creson, T., Zhang, L., Li, P., Du, F., Yuan, P., Gould, T.D., Manji, H.K., and Chen, G. (2004). Mood stabilizer valproate promotes ERK pathway-dependent cortical neuronal growth and neurogenesis. J Neurosci *24*, 6590-6599.

Harada, R., Vadnais, C., Sansregret, L., Leduy, L., Berube, G., Robert, F., and Nepveu, A. (2008). Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes. Nucleic Acids Res *36*, 189-202.

Harrison, P.J. (1997). Schizophrenia: a disorder of neurodevelopment? Current opinion in neurobiology *7*, 285-289.

Harrison, P.J. (1999). The neuropathology of schizophrenia. A critical review of the data and their interpretation. Brain *122 (Pt 4)*, 593-624.

Harrison, P.J. (2004). The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. Psychopharmacology (Berl) *174*, 151-162.

Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. Nature reviews Genetics *10*, 551-564.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L*., et al.* (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391-399.

Hayashi-Takagi, A., and Sawa, A. (2010). Disturbed synaptic connectivity in schizophrenia: convergence of genetic risk factors during neurodevelopment. Brain Res Bull *83*, 140-146.

Hebert, S.S., Horre, K., Nicolai, L., Bergmans, B., Papadopoulou, A.S., Delacourte, A., and De Strooper, B. (2009). MicroRNA regulation of Alzheimer's Amyloid precursor protein expression. Neurobiol Dis *33*, 422-428.

Heckers, S. (1997). Neuropathology of schizophrenia: cortex, thalamus, basal ganglia, and neurotransmitter-specific projection systems. Schizophr Bull *23*, 403-421.

Henthorn, P., McCarrick-Walmsley, R., and Kadesch, T. (1990). Sequence of the cDNA encoding ITF-2, a positive-acting transcription factor. Nucleic Acids Res *18*, 678.

Herbst, A., Helferich, S., Behrens, A., Göke, B., and Kolligs, F.T. (2009). The transcription factor ITF-2A induces cell cycle arrest via p21Cip1☆. Biochemical and Biophysical Research Communications *387*, 736-740.

Herbst, A., and Kolligs, F. (2008). A conserved domain in the transcription factor ITF-2B attenuates its activity. Biochemical and Biophysical Research Communications *370*, 327-331.

Hill, M.J., and Bray, N.J. (2012). Evidence that schizophrenia risk variation in the ZNF804A gene exerts its effects during fetal brain development. Am J Psychiatry *169*, 1301-1308.

Hill, M.J., Donocik, J.G., Nuamah, R.A., Mein, C.A., Sainz-Fuertes, R., and Bray, N.J. (2014). Transcriptional consequences of schizophrenia candidate miR-137 manipulation in human neural progenitor cells. Schizophr Res *153*, 225-230.

Howes, O.D., and Kapur, S. (2009). The dopamine hypothesis of schizophrenia: version III--the final common pathway. Schizophr Bull *35*, 549-562.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics *2008*, 420747.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Hurles, M.E., Dermitzakis, E.T., and Tyler-Smith, C. (2008). The functional impact of structural variation in humans. Trends Genet *24*, 238-245.

Ikeda, M., Aleksic, B., Kirov, G., Kinoshita, Y., Yamanouchi, Y., Kitajima, T., Kawashima, K., Okochi, T., Kishi, T., Zaharieva, I.*, et al.* (2010). Copy number variation in schizophrenia in the Japanese population. Biol Psychiatry *67*, 283-286.

Impagnatiello, F., Guidotti, A.R., Pesold, C., Dwivedi, Y., Caruncho, H., Pisu, M.G., Uzunov, D.P., Smalheiser, N.R., Davis, J.M., Pandey, G.N.*, et al.* (1998). A decrease of reelin expression as a putative vulnerability factor in schizophrenia. Proc Natl Acad Sci U S A *95*, 15718-15723.

Ince-Dunn, G., Hall, B.J., Hu, S.C., Ripley, B., Huganir, R.L., Olson, J.M., Tapscott, S.J., and Ghosh, A. (2006). Regulation of thalamocortical patterning and synaptic maturation by NeuroD2. Neuron *49*, 683-695.

Ingason, A., Rujescu, D., Cichon, S., Sigurdsson, E., Sigmundsson, T., Pietilainen, O.P., Buizer-Voskamp, J.E., Strengman, E., Francks, C., Muglia, P.*, et al.* (2011). Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol Psychiatry *16*, 17-25.

International Schizophrenia, C. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature *455*, 237-241.

International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748-752.

Irish Schizophrenia Genomics, C., and the Wellcome Trust Case Control, C. (2012). Genome-wide association study implicates HLA-C*01:02 as a risk factor at the major histocompatibility complex locus in schizophrenia. Biol Psychiatry *72*, 620-628.

Javitt, D.C., and Zukin, S.R. (1991). Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry *148*, 1301-1308.

Jin, C., Samuelson, L., Cui, C.B., Sun, Y., and Gerber, D.A. (2011). MAPK/ERK and Wnt/beta-Catenin pathways are synergistically involved in proliferation of Sca-1 positive hepatic progenitor cells. Biochem Biophys Res Commun *409*, 803-807.

Johansen, C.T., Wang, J., Lanktree, M.B., Cao, H., McIntyre, A.D., Ban, M.R., Martins, R.A., Kennedy, B.A., Hassell, R.G., Visser, M.E.*, et al.* (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet *42*, 684-687.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938-2939.

Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanovic, D., Geschwind, D.H., Mane, S.M., State, M.W., and Sestan, N. (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron *62*, 494-509.

Joyce, J.N., and Meador-Woodruff, J.H. (1997). Linking the family of D2 receptors to neuronal circuits in human brain: insights into schizophrenia. Neuropsychopharmacology *16*, 375-384.

Jung, G.A., Yoon, J.Y., Moon, B.S., Yang, D.H., Kim, H.Y., Lee, S.H., Bryja, V., Arenas, E., and Choi, K.Y. (2008). Valproic acid induces differentiation and inhibition of proliferation in neural progenitor cells via the beta-catenin-Ras-ERK-p21Cip/WAF1 pathway. BMC cell biology *9*, 66.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G.*, et al.* (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483-489.

Kao, C.Y., Hsu, Y.C., Liu, J.W., Lee, D.C., Chung, Y.F., and Chiu, I.M. (2013). The mood stabilizer valproate activates human FGF1 gene promoter through inhibiting HDAC and GSK-3 activities. J Neurochem *126*, 4-18.

Kim, S., Park, D.Y., Lee, D., Kim, W., Jeong, Y.H., Lee, J., Chung, S.K., Ha, H., Choi, B.H., and Kim, K.T. (2014). Vaccinia-related kinase 2 mediates accumulation of polyglutamine aggregates via negative regulation of the chaperonin TRiC. Mol Cell Biol *34*, 643-652.

Kimoto, S., Okuda, A., Toritsuka, M., Yamauchi, T., Makinodan, M., Okuda, H., Tatsumi, K., Nakamura, Y., Wanaka, A., and Kishimoto, T. (2011). Olanzapine stimulates proliferation but inhibits differentiation in rat oligodendrocyte precursor cell cultures. Progress in neuro-psychopharmacology & biological psychiatry *35*, 1950-1956.

Kirkbride, J.B., Errazuriz, A., Croudace, T.J., Morgan, C., Jackson, D., Boydell, J., Murray, R.M., and Jones, P.B. (2012). Incidence of schizophrenia and other psychoses in England, 1950-2009: a systematic review and meta-analyses. PLoS One *7*, e31660.

Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., International Schizophrenia, C., Wellcome Trust Case Control, C., Craddock, N., Owen, M.J.*, et al.* (2009a). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Hum Mol Genet *18*, 1497-1503.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L.*, et al.* (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry *17*, 142-153.

Kirov, G., Rujescu, D., Ingason, A., Collier, D.A., O'Donovan, M.C., and Owen, M.J. (2009b). Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophr Bull *35*, 851-854.

Kolligs, F.T., Nieman, M.T., Winer, I., Hu, G., Van Mater, D., Feng, Y., Smith, I.M., Wu, R., Zhai, Y., Cho, K.R.*, et al.* (2002). ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. Cancer Cell *1*, 145-155.

Kuipers, H., Schnorfeil, F.M., and Brocker, T. (2010). Differentially expressed microRNAs regulate plasmacytoid vs. conventional dendritic cell development. Mol Immunol *48*, 333-340.

Kuot, A., Hewitt, A.W., Griggs, K., Klebe, S., Mills, R., Jhanji, V., Craig, J.E., Sharma, S., and Burdon, K.P. (2012). Association of TCF4 and CLU polymorphisms with Fuchs'

endothelial dystrophy and implication of CLU and TGFBI proteins in the disease process. Eur J Hum Genet *20*, 632-638.

Kurian, S.M., Le-Niculescu, H., Patel, S.D., Bertram, D., Davis, J., Dike, C., Yehyawi, N., Lysaker, P., Dustin, J., Caligiuri, M.*, et al.* (2011). Identification of blood biomarkers for psychosis using convergent functional genomics. Mol Psychiatry *16*, 37-58.

Kurihara, Y., Suzuki, T., Sakaue, M., Murayama, O., Miyazaki, Y., Onuki, A., Aoki, T., Saito, M., Fujii, Y., Hisasue, M.*, et al.* (2014). Valproic Acid, a Histone Deacetylase Inhibitor, Decreases Proliferation of and Induces Specific Neurogenic Differentiation of Canine Adipose Tissue-Derived Stem Cells. Journal of Veterinary Medical Science *76*, 15-23.

Kwon, E., Wang, W., and Tsai, L.H. (2013). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol Psychiatry *18*, 11-12.

Kyosseva, S.V., Elbein, A.D., Griffin, W.S., Mrak, R.E., Lyon, M., and Karson, C.N. (1999). Mitogen-activated protein kinases in schizophrenia. Biol Psychiatry *46*, 689-696.

la Cour, T., Kiemer, L., Molgaard, A., Gupta, R., Skriver, K., and Brunak, S. (2004). Analysis and prediction of leucine-rich nuclear export signals. Protein Eng Des Sel *17*, 527-536.

Lahti, A.C., Koffel, B., LaPorte, D., and Tamminga, C.A. (1995). Subanesthetic doses of ketamine stimulate psychosis in schizophrenia. Neuropsychopharmacology *13*, 9-19.

Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373-379.

Lander, E.S., and Schork, N.J. (1994). Genetic dissection of complex traits. Science *265*, 2037-2048.

Lang, G.K., and Naumann, G.O. (1987). The frequency of corneal dystrophies requiring keratoplasty in Europe and the U.S.A. Cornea *6*, 209-211.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Langlands, K., Yin, X., Anand, G., and Prochownik, E.V. (1997). Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. J Biol Chem *272*, 19785-19793.

Lappalainen, T., Montgomery, S.B., Nica, A.C., and Dermitzakis, E.T. (2011). Epistatic selection between coding and regulatory variation in human evolution and disease. Am J Hum Genet *89*, 459-463.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R.*, et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Laruelle, M. (1998). Imaging dopamine transmission in schizophrenia. A review and meta-analysis. The quarterly journal of nuclear medicine : official publication of the Italian Association of Nuclear Medicine *42*, 211-221.

Laruelle, M., Abi-Dargham, A., van Dyck, C.H., Gil, R., D'Souza, C.D., Erdos, J., McCance, E., Rosenblatt, W., Fingado, C., Zoghbi, S.S.*, et al.* (1996). Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Natl Acad Sci U S A *93*, 9235-9240.

Laterza, O.F., Lim, L., Garrett-Engele, P.W., Vlasakova, K., Muniappa, N., Tanaka, W.K., Johnson, J.M., Sina, J.F., Fare, T.L., Sistare, F.D.*, et al.* (2009). Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clin Chem *55*, 1977-1983.

Lavon, I., Zrihan, D., Granit, A., Einstein, O., Fainstein, N., Cohen, M.A., Zelikovitch, B., Shoshan, Y., Spektor, S., Reubinoff, B.E.*, et al.* (2010). Gliomas display a microRNA expression profile reminiscent of neural precursor cells. Neuro Oncol *12*, 422-433.

Lawrie, S.M. (2001). Neurodevelopmental indices and the development of psychotic symptoms in subjects at high risk of schizophrenia. The British Journal of Psychiatry *178*, 524-530.

Lazorchak, A., Jones, M.E., and Zhuang, Y. (2005). New insights into E-protein function in lymphocyte development. Trends Immunol *26*, 334-338.

Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., Zhang, N., Mowry, B.J., Olincy, A., Amin, F.*, et al.* (2011). Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry *168*, 302-316.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, C., Levinson, D., Wise, L., Delisi, L., Straub, R., Hovatta, I., Williams, N., Schwab, S., Pulver, A., and Faraone, S. (2003a). Genome Scan Meta-Analysis of Schizophrenia and Bipolar Disorder, Part II: Schizophrenia. The American Journal of Human Genetics *73*, 34-48.

Lewis, D.A., Glantz, L.A., Pierri, J.N., and Sweet, R.A. (2003b). Altered cortical glutamate neurotransmission in schizophrenia: evidence from morphological studies of pyramidal neurons. Ann N Y Acad Sci *1003*, 102-112.

Lewis, D.A., and Gonzalez-Burgos, G. (2008). Neuroplasticity of neocortical circuits in schizophrenia. Neuropsychopharmacology *33*, 141-165.

Lewis, D.A., and Levitt, P. (2002). Schizophrenia as a disorder of neurodevelopment. Annual review of neuroscience *25*, 409-432.

Lewis, D.A., and Lieberman, J.A. (2000). Catching up on schizophrenia: natural history and neurobiology. Neuron *28*, 325-334.

Li, H. (2011). Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics *27*, 718-719.

Li, L.Y., Liu, M.Y., Shih, H.M., Tsai, C.H., and Chen, J.Y. (2006). Human cellular protein VRK2 interacts specifically with Epstein-Barr virus BHRF1, a homologue of Bcl-2, and enhances cell survival. J Gen Virol *87*, 2869-2878.

Li, M., Wang, Y., Zheng, X.B., Ikeda, M., Iwata, N., Luo, X.J., Chong, S.A., Lee, J., Rietschel, M., Zhang, F.*, et al.* (2012). Meta-analysis and brain imaging data support the involvement of VRK2 (rs2312147) in schizophrenia susceptibility. Schizophr Res *142*, 200-205.

Li, Y.J., Minear, M.A., Rimmler, J., Zhao, B., Balajonda, E., Hauser, M.A., Allingham, R.R., Eghrari, A.O., Riazuddin, S.A., Katsanis, N.*, et al.* (2011). Replication of TCF4 through association and linkage studies in late-onset Fuchs endothelial corneal dystrophy. PLoS One *6*, e18044.

Liang, J.S., Shimojima, K., Ohno, K., Sugiura, C., Une, Y., and Yamamoto, T. (2009). A newly recognised microdeletion syndrome of 2p15-16.1 manifesting moderate developmental delay, autistic behaviour, short stature, microcephaly, and dysmorphic features: a new patient with 3.2 Mb deletion. J Med Genet *46*, 645-647.

Lichtermann, D., Karbe, E., and Maier, W. (2000). The genetic epidemiology of schizophrenia and of schizophrenia spectrum disorders. Eur Arch Psychiatry Clin Neurosci *250*, 304-310.

Lieberman, J.A., and Koreen, A.R. (1993). Neurochemistry and neuroendocrinology of schizophrenia: a selective review. Schizophr Bull *19*, 371-429.

Lin, C.H., Tapscott, S.J., and Olson, J.M. (2006). Congenital hypothyroidism (cretinism) in neuroD2-deficient mice. Mol Cell Biol *26*, 4311-4315.

Littlewood, T.D., Hancock, D.C., Danielian, P.S., Parker, M.G., and Evan, G.I. (1995). A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. Nucleic Acids Res *23*, 1686-1690.

Liu, D.Z., Tian, Y., Ander, B.P., Xu, H., Stamova, B.S., Zhan, X., Turner, R.J., Jickling, G., and Sharp, F.R. (2010). Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. J Cereb Blood Flow Metab *30*, 92-101.

Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M.*, et al.* (2014). DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Molecular autism *5*, 22.

Liu, P.Y., Hsieh, T.Y., Liu, S.T., Chang, Y.L., Lin, W.S., Wang, W.M., and Huang, S.M. (2011). Zac1, an Sp1-like protein, regulates human p21(WAF1/Cip1) gene expression in HeLa cells. Experimental cell research *317*, 2925-2937.

Liu, Y., Ray, S.K., Yang, X.Q., Luntz-Leybman, V., and Chiu, I.M. (1998). A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. J Biol Chem *273*, 19269-19276.

Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., and Cui, Q. (2008). An analysis of human microRNA and disease associations. PLoS ONE *3*, e3420.

Lu, R., Qu, Y., Ge, J., Zhang, L., Su, Z., Pflugfelder, S.C., and Li, D.Q. (2012). Transcription factor TCF4 maintains the properties of human corneal epithelial stem cells. Stem Cells *30*, 753-761.

Lukk, M., Kapushesky, M., Nikkila, J., Parkinson, H., Goncalves, A., Huber, W., Ukkonen, E., and Brazma, A. (2010). A global map of human gene expression. Nat Biotechnol *28*, 322-324.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci U S A *104 Suppl 1*, 8597-8604.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A.*, et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508*, 469-476.

Magri, C., Sacchetti, E., Traversa, M., Valsecchi, P., Gardella, R., Bonvicini, C., Minelli, A., Gennarelli, M., and Barlati, S. (2010). New copy number variations in schizophrenia. PLoS One *5*, e13422.

Malaterre, J., Ramsay, R.G., and Mantamadiotis, T. (2007). Wnt-Frizzled signalling and the many paths to neural development and adult brain homeostasis. Frontiers in bioscience : a journal and virtual library *12*, 492-506.

Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell *148*, 1223-1241.

Mao, Y., Ge, X., Frank, C.L., Madison, J.M., Koehler, A.N., Doud, M.K., Tassa, C., Berry, E.M., Soda, T., Singh, K.K.*, et al.* (2009). Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. Cell *136*, 1017-1031.

Maycox, P.R., Kelly, F., Taylor, A., Bates, S., Reid, J., Logendra, R., Barnes, M.R., Larminie, C., Jones, N., Lennon, M.*, et al.* (2009). Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Mol Psychiatry *14*, 1083-1094.

McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O.*, et al.* (2009). Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet *41*, 1223-1227.

McGrath, J., Saha, S., Chant, D., and Welham, J. (2008). Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiologic reviews *30*, 67-76.

McGrath, J.J., Burne, T.H., Feron, F., Mackay-Sim, A., and Eyles, D.W. (2010). Developmental vitamin D deficiency and risk of schizophrenia: a 10-year update. Schizophr Bull *36*, 1073-1078.

McGuffin, P., Tandon, K., and Corsico, A. (2003). Linkage and association studies of schizophrenia. Curr Psychiatry Rep *5*, 121-127.

McInnis, M.G., Swift-Scanlanl, T., Mahoney, A.T., Vincent, J., Verheyen, G., Lan, T.H., Oruc, L., Riess, O., Van Broeckhoven, C., Chen, H.*, et al.* (2000). Allelic distribution of CTG18.1 in Caucasian populations: association studies in bipolar disorder, schizophrenia, and ataxia. Mol Psychiatry *5*, 439-442.

Meek, K.M., Leonard, D.W., Connon, C.J., Dennis, S., and Khan, S. (2003). Transparency, swelling and scarring in the corneal stroma. Eye (Lond) *17*, 927-936.

Melom, J.E., and Littleton, J.T. (2011). Synapse development in health and disease. Curr Opin Genet Dev *21*, 256-261.

Merz, K., Herold, S., and Lie, D.C. (2011). CREB in adult neurogenesis--master and partner in the development of adult-born neurons? Eur J Neurosci *33*, 1078-1086.

Miyaoka, T., Seno, H., and Ishino, H. (1999). Increased expression of Wnt-1 in schizophrenic brains. Schizophr Res *38*, 1-6.

Monsalve, D.M., Merced, T., Fernandez, I.F., Blanco, S., Vazquez-Cedeira, M., and Lazo, P.A. (2013). Human VRK2 modulates apoptosis by interaction with Bcl-xL and regulation of BAX gene expression. Cell death & disease *4*, e513.

Moreno-De-Luca, D., Consortium, S., Mulle, J.G., Simons Simplex Collection Genetics, C., Kaminsky, E.B., Sanders, S.J., GeneStar, Myers, S.M., Adam, M.P., Pakula, A.T.*, et al.* (2010). Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet *87*, 618-630.

Mudge, J., Miller, N.A., Khrebtukova, I., Lindquist, I.E., May, G.D., Huntley, J.J., Luo, S., Zhang, L., van Velkinburgh, J.C., Farmer, A.D.*, et al.* (2008). Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. PLoS ONE *3*, e3625.

Muir, T., Sadler-Riggleman, I., Stevens, J.D., and Skinner, M.K. (2006). Role of the basic helix-loop-helix protein ITF2 in the hormonal regulation of Sertoli cell differentiation. Molecular Reproduction and Development *73*, 491-500.

Muir, W.J., Pickard, B.S., and Blackwood, D.H. (2008). Disrupted-in-Schizophrenia-1. Curr Psychiatry Rep *10*, 140-147.

Mulle, J.G., Dodd, A.F., McGrath, J.A., Wolyniec, P.S., Mitchell, A.A., Shetty, A.C., Sobreira, N.L., Valle, D., Rudd, M.K., Satten, G.*, et al.* (2010). Microdeletions of 3q29 confer high risk for schizophrenia. Am J Hum Genet *87*, 229-236.

Murray, R.M., and Lewis, S.W. (1987). Is schizophrenia a neurodevelopmental disorder? British medical journal *295*, 681-682.

Murre, C. (2005). Helix-loop-helix proteins and lymphocyte development. Nature Immunology *6*, 1079-1086.

Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B.*, et al.* (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell *58*, 537-544.

Muzio, L., and Consalez, G.G. (2013). Modeling human brain development with cerebral organoids. Stem cell research & therapy *4*, 154.

Nagasawa, M., Schmidlin, H., Hazekamp, M.G., Schotte, R., and Blom, B. (2008). Development of human plasmacytoid dendritic cells depends on the combined action of the basic helix-loop-helix factor E2-2 and the Ets factor Spi-B. European Journal of Immunology *38*, 2389-2400.

Narayan, S., Tang, B., Head, S.R., Gilmartin, T.J., Sutcliffe, J.G., Dean, B., and Thomas, E.A. (2008). Molecular profiles of schizophrenia in the CNS at different stages of illness. Brain Res *1239*, 235-248.

Nass, D., Rosenwald, S., Meiri, E., Gilad, S., Tabibian-Keissar, H., Schlosberg, A., Kuker, H., Sion-Vardy, N., Tobar, A., Kharenko, O.*, et al.* (2009). MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. Brain Pathol *19*, 375-383.

Navarrete, K., Pedroso, I., De Jong, S., Stefansson, H., Steinberg, S., Stefansson, K., Ophoff, R.A., Schalkwyk, L.C., and Collier, D.A. (2013). TCF4 (e2-2; ITF2): a schizophrenia-associated gene with pleiotropic effects on human disease. Am J Med Genet B Neuropsychiatr Genet *162b*, 1-16.

Need, A.C., Ge, D., Weale, M.E., Maia, J., Feng, S., Heinzen, E.L., Shianna, K.V., Yoon, W., Kasperaviciute, D., Gennarelli, M.*, et al.* (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS Genet *5*, e1000373.

Nezu, J., Oku, A., Jones, M.H., and Shimane, M. (1997). Identification of two novel human putative serine/threonine kinases, VRK1 and VRK2, with structural similarity to vaccinia virus B1R kinase. Genomics *45*, 327-331.

Ng, M.Y., Levinson, D.F., Faraone, S.V., Suarez, B.K., DeLisi, L.E., Arinami, T., Riley, B., Paunio, T., Pulver, A.E., Irmansyah*, et al.* (2009). Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol Psychiatry *14*, 774-785.

Nica, A.C., and Dermitzakis, E.T. (2008). Using gene expression to investigate the genetic basis of complex disorders. Hum Mol Genet *17*, R129-134.

Nichols, R.J., and Traktman, P. (2004). Characterization of three paralogous members of the Mammalian vaccinia related kinase family. J Biol Chem *279*, 7934-7946.

Nichols, R.J., Wiebe, M.S., and Traktman, P. (2006). The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. Mol Biol Cell *17*, 2451-2464.

Niu, J., Mei, F., Li, N., Wang, H., Li, X., Kong, J., and Xiao, L. (2010). Haloperidol promotes proliferation but inhibits differentiation in rat oligodendrocyte progenitor cell cultures. Biochemistry and cell biology = Biochimie et biologie cellulaire *88*, 611-620.

Niznikiewicz, M.A.K., M. Shenton M. (2003). Recent structural and functional imaging findings in schizophrenia. Current opinion in psychiatry *16*, 123–147.

Noda, T., Kawamura, R., Funabashi, H., Mie, M., and Kobatake, E. (2006). Transduction of NeuroD2 protein induced neural cell differentiation. J Biotechnol *126*, 230-236.

Northcott, P.A., Fernandez, L.A., Hagan, J.P., Ellison, D.W., Grajkowska, W., Gillespie, Y., Grundy, R., Van Meter, T., Rutka, J.T., Croce, C.M.*, et al.* (2009). The miR-17/92 polycistron is up-regulated in sonic hedgehog-driven medulloblastomas and induced by

N-myc in sonic hedgehog-treated cerebellar neural precursors. Cancer Res *69*, 3249-3255.

O'Dushlaine, C., Kenny, E., Heron, E., Donohoe, G., Gill, M., Morris, D., International Schizophrenia, C., and Corvin, A. (2011). Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Mol Psychiatry *16*, 286-292.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. Nature Neuroscience *11*, 1271-1282.

Olney, J.W., Newcomer, J.W., and Farber, N.B. (1999). NMDA receptor hypofunction model of schizophrenia. Journal of psychiatric research *33*, 523-533.

Olson, J.M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J., and Tapscott, S.J. (2001). NeuroD2 is necessary for development and survival of central nervous system neurons. Dev Biol *234*, 174-187.

Oni-Orisan, A., Kristiansen, L.V., Haroutunian, V., Meador-Woodruff, J.H., and McCullumsmith, R.E. (2008). Altered vesicular glutamate transporter expression in the anterior cingulate cortex in schizophrenia. Biol Psychiatry *63*, 766-775.

Orrico, A., Galli, L., Zappella, M., Lam, C.W., Bonifacio, S., Torricelli, F., and Hayek, G. (2001). Possible case of Pitt-Hopkins syndrome in sibs. Am J Med Genet *103*, 157-159.

Othman, A., Frim, D.M., Polak, P., Vujicic, S., Arnason, B.G., and Boullerne, A.I. (2011). Olig1 is expressed in human oligodendrocytes during maturation and regeneration. Glia *59*, 914-926.

Owen, M.J., Williams, H.J., and O'Donovan, M.C. (2009). Schizophrenia genetics: advancing on two fronts. Curr Opin Genet Dev *19*, 266-270.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nature reviews Genetics *12*, 87-98.

P.H., C. (1958). Amphetamine Psychosis. Oxford University Press, London.

Pagliuca, A., Gallo, P., De Luca, P., and Lania, L. (2000). Class A helix-loop-helix proteins are positive regulators of several cyclin-dependent kinase inhibitors' promoter activity and negatively affect cell growth. Cancer Res *60*, 1376-1382.

Pagotto, U., Arzberger, T., Ciani, E., Lezoualc'h, F., Pilon, C., Journot, L., Spengler, D., and Stalla, G.K. (1999). Inhibition of Zac1, a new gene differentially expressed in the anterior pituitary, increases cell proliferation. Endocrinology *140*, 987-996.

Panman, L., Andersson, E., Alekseenko, Z., Hedlund, E., Kee, N., Mong, J., Uhde, C.W., Deng, Q., Sandberg, R., Stanton, L.W.*, et al.* (2011). Transcription factor-induced lineage selection of stem-cell-derived neural progenitor cells. Cell stem cell *8*, 663-675.

Parrinello, S., Lin, C.Q., Murata, K., Itahana, Y., Singh, J., Krtolica, A., Campisi, J., and Desprez, P.Y. (2001). Id-1, ITF-2, and Id-2 comprise a network of helix-loop-helix proteins that regulate mammary epithelial cell proliferation, differentiation, and apoptosis. J Biol Chem *276*, 39213-39219.

Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. Nature reviews Genetics *11*, 533-538.

Pawel, K., Hauser, J., Skibinska, M., Szczepankiewicz, A., Dmitrzak-Weglarz, M., Gorzkowska, K., Pawlak, J., and Czerski, P.M. (2010). [Family based association study of DRD1, DRD2, DRD3, DRD4, DAT, COMT gene polymorphism in schizophrenia]. Psychiatria polska *44*, 405-413.

Peippo, M.M., Simola, K.O., Valanne, L.K., Larsen, A.T., Kahkonen, M., Auranen, M.P., and Ignatius, J. (2006). Pitt-Hopkins syndrome in two patients and further definition of the phenotype. Clin Dysmorphol *15*, 47-54.

Pemberton, L.F., and Paschal, B.M. (2005). Mechanisms of Receptor-Mediated Nuclear Import and Nuclear Export. Traffic *6*, 187-198.

Perkins, D.O., Jeffries, C.D., Jarskog, L.F., Thomson, J.M., Woods, K., Newman, M.A., Parker, J.S., Jin, J., and Hammond, S.M. (2007). microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. Genome Biol *8*, R27.

Persson, P., Jogi, A., Grynfeld, A., Pahlman, S., and Axelson, H. (2000). HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. Biochem Biophys Res Commun *274*, 22-31.

Petropoulos, H., and Skerjanc, I.S. (2000). Analysis of the inhibition of MyoD activity by ITF-2B and full-length E12/E47. J Biol Chem *275*, 25095-25101.

Pickard, B.S., Malloy, M.P., Clark, L., Lehellard, S., Ewald, H.L., Mors, O., Porteous, D.J., Blackwood, D.H., and Muir, W.J. (2005). Candidate psychiatric illness genes identified in patients with pericentric inversions of chromosome 18. Psychiatr Genet *15*, 37-44.

Pidsley, R., and Mill, J. (2011). Epigenetic studies of psychosis: current findings, methodological approaches, and implications for postmortem research. Biol Psychiatry *69*, 146-156.

Pierson, J., Hostager, B., Fan, R., and Vibhakar, R. (2008). Regulation of cyclin dependent kinase 6 by microRNA 124 in medulloblastoma. J Neurooncol *90*, 1-7.

Pitt, D., and Hopkins, I. (1978). A syndrome of mental retardation, wide mouth and intermittent overbreathing. Aust Paediatr J *14*, 182-184.

Poethig, R.S. (2009). Small RNAs and developmental timing in plants. Curr Opin Genet Dev *19*, 374-378.

Pollock, K., Stroemer, P., Patel, S., Stevanato, L., Hope, A., Miljan, E., Dong, Z., Hodges, H., Price, J., and Sinden, J. (2006). A conditionally immortal clonal stem cell line from human cortical neuroepithelium for the treatment of ischemic stroke. Experimental Neurology *199*, 143-155.

Pscherer, A., Dorflinger, U., Kirfel, J., Gawlas, K., Ruschoff, J., Buettner, R., and Schule, R. (1996). The helix-loop-helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin receptor II gene. EMBO J *15*, 6680-6690.

Psychiatric, G.C.B.D.W.G. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet *43*, 977-983.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kahler, A.*, et al.* (2014). A polygenic burden of rare disruptive mutations in schizophrenia. Nature *506*, 185-190.

Quednow, B.B., Brinkmeyer, J., Mobascher, A., Nothnagel, M., Musso, F., Grunder, G., Savary, N., Petrovsky, N., Frommann, I., Lennertz, L.*, et al.* (2012). Schizophrenia risk polymorphisms in the TCF4 gene interact with smoking in the modulation of auditory sensory gating. Proc Natl Acad Sci U S A *109*, 6271-6276.

Quednow, B.B., Ettinger, U., Mossner, R., Rujescu, D., Giegling, I., Collier, D.A., Schmechtig, A., Kuhn, K.U., Moller, H.J., Maier, W.*, et al.* (2011). The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. J Neurosci *31*, 6684-6691.

Quong, M.W., Romanow, W.J., and Murre, C. (2002). E Protein Function in Lymphocyte Development. Annual Review of Immunology *20*, 301-322.

Ragland, J.D., Yoon, J., Minzenberg, M.J., and Carter, C.S. (2007). Neuroimaging of cognitive disability in schizophrenia: search for a pathophysiological mechanism. Int Rev Psychiatry *19*, 417-427.

Rajcan-Separovic, E., Harvard, C., Liu, X., McGillivray, B., Hall, J.G., Qiao, Y., Hurlburt, J., Hildebrand, J., Mickelson, E.C., Holden, J.J.*, et al.* (2007). Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. J Med Genet *44*, 269-276.

Ravanpay, A.C., and Olson, J.M. (2008). E protein dosage influences brain development more than family member identity. Journal of Neuroscience Research *86*, 1472-1481.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N.*, et al.* (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell *140*, 744-752.

Reif, A., Fritzen, S., Finger, M., Strobel, A., Lauer, M., Schmitt, A., and Lesch, K.P. (2006). Neural stem cell proliferation is decreased in schizophrenia, but not in depression. Mol Psychiatry *11*, 514-522.

Reif, A., Schmitt, A., Fritzen, S., and Lesch, K.P. (2007). Neurogenesis and schizophrenia: dividing neurons in a divided mind? Eur Arch Psychiatry Clin Neurosci *257*, 290-299.

Richards, A.L., Jones, L., Moskvina, V., Kirov, G., Gejman, P.V., Levinson, D.F., Sanders, A.R., Molecular Genetics of Schizophrenia, C., International Schizophrenia, C., Purcell, S.*, et al.* (2012). Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. Mol Psychiatry *17*, 193-201.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kahler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M.*, et al.* (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet *45*, 1150-1159.

Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A.*, et al.* (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet *43*, 969-976.

Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science *273*, 1516-1517.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A.*, et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods *4*, 651-657.

Ross, S.E., Greenberg, M.E., and Stiles, C.D. (2003). Basic helix-loop-helix factors in cortical development. Neuron *39*, 13-25.

Rossbach, M. (2011). Non-Coding RNAs in Neural Networks, REST-Assured. Front Genet *2*, 8.

Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M.B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol *27*, 66-75.

Rujescu, D., and Collier, D.A. (2009). Dissecting the many genetic faces of schizophrenia. Epidemiol Psichiatr Soc *18*, 91-95.

Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E.*, et al.* (2009). Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet *18*, 988-996.

Saarikettu, J., Sveshnikova, N., and Grundstrom, T. (2004). Calcium/calmodulin inhibition of transcriptional activity of E-proteins by prevention of their binding to DNA. J Biol Chem *279*, 41004-41011.

Saba, R., Goodman, C.D., Huzarewich, R.L., Robertson, C., and Booth, S.A. (2008). A miRNA signature of prion induced neurodegeneration. PLoS ONE *3*, e3652.

Saha, S., Chant, D., Welham, J., and McGrath, J. (2005). A systematic review of the prevalence of schizophrenia. PLoS medicine *2*, e141.

Sanz-Garcia, M., Lopez-Sanchez, I., and Lazo, P.A. (2008). Proteomics identification of nuclear Ran GTPase as an inhibitor of human VRK1 and VRK2 (vaccinia-related kinase) activities. Mol Cell Proteomics *7*, 2199-2214.

Sanz-Garcia, M., Vazquez-Cedeira, M., Kellerman, E., Renbaum, P., Levy-Lahad, E., and Lazo, P.A. (2011). Substrate profiling of human vaccinia-related kinases identifies coilin, a Cajal body nuclear protein, as a phosphorylation target with neurological implications. Journal of proteomics *75*, 548-560.

Schizophrenia Psychiatric Genome-Wide Association Study, C. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet *43*, 969-976.

Schmidt-Edelkraut, U., Daniel, G., Hoffmann, A., and Spengler, D. (2014). Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. Mol Cell Biol *34*, 1020-1030.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675.

Schreiber, M., Kolbus, A., Piu, F., Szabowski, A., Mohle-Steinlein, U., Tian, J., Karin, M., Angel, P., and Wagner, E.F. (1999). Control of cell cycle progression by c-Jun is p53 dependent. Genes Dev *13*, 607-619.

Schwarz, C., Volz, A., Li, C., and Leucht, S. (2008). Valproate for schizophrenia. Cochrane Database Syst Rev, CD004028.

Seeman, P., Chau-Wong, M., Tedesco, J., and Wong, K. (1975). Brain receptors for antipsychotic drugs and dopamine: direct binding assays. Proc Natl Acad Sci U S A *72*, 4376-4380.

Selemon, L.D., and Goldman-Rakic, P.S. (1999). The reduced neuropil hypothesis: a circuit based model of schizophrenia. Biol Psychiatry *45*, 17-25.

Sepp, M., Kannike, K., Eesmaa, A., Urb, M., and Timmusk, T. (2011). Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. PLoS ONE *6*, e22138.

Sethupathy, P., Borel, C., Gagnebin, M., Grant, G.R., Deutsch, S., Elton, T.S., Hatzigeorgiou, A.G., and Antonarakis, S.E. (2007). Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3'

untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. Am J Hum Genet *81*, 405-413.

Sevilla, A., Santos, C.R., Barcia, R., Vega, F.M., and Lazo, P.A. (2004). c-Jun phosphorylation by the human vaccinia-related kinase 1 (VRK1) and its cooperation with the N-terminal kinase of c-Jun (JNK). Oncogene *23*, 8950-8958.

Shepherd, A.M., Laurens, K.R., Matheson, S.L., Carr, V.J., and Green, M.J. (2012). Systematic meta-review and quality assessment of the structural brain alterations in schizophrenia. Neuroscience and biobehavioral reviews *36*, 1342-1356.

Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J.*, et al.* (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753-757.

Shi, Y., Li, Z., Xu, Q., Wang, T., Li, T., Shen, J., Zhang, F., Chen, J., Zhou, G., Ji, W.*, et al.* (2011). Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. Nat Genet *43*, 1224-1227.

Shih, R.A., Belmonte, P.L., and Zandi, P.P. (2004). A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. Int Rev Psychiatry *16*, 260-283.

Silber, J., Lim, D.A., Petritsch, C., Persson, A.I., Maunakea, A.K., Yu, M., Vandenberg, S.R., Ginzinger, D.G., James, C.D., Costello, J.F.*, et al.* (2008). miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. BMC Medicine *6*, 14.

Singh, H.A. (1993). Mental retardation, macrostomia and hyperpnoea syndrome. J Paediatr Child Health *29*, 156-157.

Sintoni, S., Kurtys, E., Scandaglia, M., Contestabile, A., and Monti, B. (2013). Chronic valproic acid administration impairs contextual memory and dysregulates hippocampal GSK-3beta in rats. Pharmacology, biochemistry, and behavior *106*, 8-15.

Skerjanc, I.S., Truong, J., Filion, P., and McBurney, M.W. (1996). A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. J Biol Chem *271*, 3555-3561.

Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.-Q., Luo, Y., Peng, J., Bordey, A.*, et al.* (2010). MicroRNA miR-137 Regulates Neuronal Maturation by Targeting Ubiquitin Ligase Mind Bomb-1. Stem Cells *28*, 1060-1070.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Sobrado, V.R., Moreno-Bueno, G., Cubillo, E., Holt, L.J., Nieto, M.A., Portillo, F., and Cano, A. (2009). The class I bHLH factors E2-2A and E2-2B regulate EMT. J Cell Sci *122*, 1014-1024.

Spengler, D., Villalba, M., Hoffmann, A., Pantaloni, C., Houssami, S., Bockaert, J., and Journot, L. (1997). Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. EMBO J *16*, 2814-2825.

St Clair, D., Blackwood, D., Muir, W., Carothers, A., Walker, M., Spowart, G., Gosden, C., and Evans, H.J. (1990). Association within a family of a balanced autosomal translocation with major mental illness. Lancet *336*, 13-16.

St Clair, D., Xu, M., Wang, P., Yu, Y., Fang, Y., Zhang, F., Zheng, X., Gu, N., Feng, G., Sham, P.*, et al.* (2005). Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961. JAMA *294*, 557-562.

Stamatoyannopoulos, J.A. (2004). The genomics of gene expression. Genomics *84*, 449-457.

Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B.*, et al.* (2009). Common variants conferring risk of schizophrenia. Nature.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E.*, et al.* (2008). Large recurrent microdeletions associated with schizophrenia. Nature *455*, 232-236.

Stefansson, H., Sigurdsson, E., Steinthorsdottir, V., Bjornsdottir, S., Sigmundsson, T., Ghosh, S., Brynjolfsson, J., Gunnarsdottir, S., Ivarsson, O., Chou, T.T.*, et al.* (2002). Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet *71*, 877-892.

Steinberg, S., de Jong, S., Irish Schizophrenia Genomics, C., Andreassen, O.A., Werge, T., Borglum, A.D., Mors, O., Mortensen, P.B., Gustafsson, O., Costas, J.*, et al.* (2011). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Hum Mol Genet *20*, 4076-4081.

Stephan, K.E., Friston, K.J., and Frith, C.D. (2009). Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bull *35*, 509-527.

Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D.*, et al.* (2007). Population genomics of human gene expression. Nat Genet *39*, 1217-1224.

Sun, M.K., Hongpaisan, J., Lim, C., and Alkon, D.L. (2014). Bryostatin-1 restores hippocampal synapses and spatial learning and memory in adult fragile X mice. The Journal of pharmacology and experimental therapeutics.

Sutton, L.P., and Rushlow, W.J. (2011). The effects of neuropsychiatric drugs on glycogen synthase kinase-3 signaling. Neuroscience *199*, 116-124.

Sweatt, J.D. (2013). Pitt-Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. Experimental & molecular medicine *45*, e21.

Talebizadeh, Z., Butler, M.G., and Theodoro, M.F. (2008). Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. Autism Res *1*, 240-250.

Tamminga, C.A., and Holcomb, H.H. (2005). Phenotype of schizophrenia: a review and formulation. Mol Psychiatry *10*, 27-39.

Tebbenkamp, A.T., Willsey, A.J., State, M.W., and Sestan, N. (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. Current opinion in neurology *27*, 149-156.

Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol *7 Suppl 1*, S12 11-14.

Tost, H., and Meyer-Lindenberg, A. (2011). Dopamine-glutamate interactions: a neural convergence mechanism of common schizophrenia risk variants. Biol Psychiatry *69*, 912-913.

Umeda-Yano, S., Hashimoto, R., Yamamori, H., Weickert, C.S., Yasuda, Y., Ohi, K., Fujimoto, M., Ito, A., and Takeda, M. (2014). Expression analysis of the genes identified in GWAS of the postmortem brain tissues from patients with schizophrenia. Neurosci Lett.

Uziel, T., Karginov, F.V., Xie, S., Parker, J.S., Wang, Y.D., Gajjar, A., He, L., Ellison, D., Gilbertson, R.J., Hannon, G.*, et al.* (2009). The miR-17~92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. Proc Natl Acad Sci U S A *106*, 2812-2817.

Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R.*, et al.* (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature *471*, 499-503.

Valente, T., Junyent, F., and Auladell, C. (2005). Zac1 is expressed in progenitor/stem cells of the neuroectoderm and mesoderm during embryogenesis: differential phenotype of the Zac1-expressing cells during development. Developmental dynamics : an official publication of the American Association of Anatomists *233*, 667-679.

Van Balkom, I.D., Vuijk, P.J., Franssens, M., Hoek, H.W., and Hennekam, R.C. (2012). Development, cognition, and behaviour in Pitt-Hopkins syndrome. Developmental medicine and child neurology *54*, 925-931.

van Os, J., and Kapur, S. (2009). Schizophrenia. Lancet *374*, 635-645.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol *3*, RESEARCH0034.

Vazquez-Cedeira, M., and Lazo, P.A. (2012). Human VRK2 (vaccinia-related kinase 2) modulates tumor cell invasion by hyperactivation of NFAT1 and expression of cyclooxygenase-2. J Biol Chem *287*, 42739-42750.

Veyrieras, J.B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M., and Pritchard, J.K. (2008). High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet *4*, e1000214.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92.

Vleugel, M.M., Greijer, A.E., Bos, R., van der Wall, E., and van Diest, P.J. (2006). c-Jun activation is associated with proliferation and angiogenesis in invasive breast cancer. Human pathology *37*, 668-674.

Volk, D.W., Austin, M.C., Pierri, J.N., Sampson, A.R., and Lewis, D.A. (2000). Decreased glutamic acid decarboxylase67 messenger RNA expression in a subset of prefrontal cortical gamma-aminobutyric acid neurons in subjects with schizophrenia. Arch Gen Psychiatry *57*, 237-245.

Vostrikov, V.M., Uranova, N.A., and Orlovskaya, D.D. (2007). Deficit of perineuronal oligodendrocytes in the prefrontal cortex in schizophrenia and mood disorders. Schizophr Res *94*, 273-280.

Walker, E.F., Lewine, R.R., and Neumann, C. (1996). Childhood behavioral characteristics and adult brain morphology in schizophrenia. Schizophr Res *22*, 93-101.

Walker, R.M., Christoforou, A., Thomson, P.A., McGhee, K.A., Maclean, A., Muhleisen, T.W., Strohmaier, J., Nieratschker, V., Nothen, M.M., Rietschel, M.*, et al.* (2010). Association analysis of Neuregulin 1 candidate regions in schizophrenia and bipolar disorder. Neurosci Lett *478*, 9-13.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A.*, et al.* (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science *320*, 539-543.

Wang, L., Lockstone, H.E., Guest, P.C., Levin, Y., Palotas, A., Pietsch, S., Schwarz, E., Rahmoune, H., Harris, L.W., Ma, D.*, et al.* (2010). Expression profiling of fibroblasts identifies cell cycle abnormalities in schizophrenia. J Proteome Res *9*, 521-527.

Wang, V.Y., Rose, M.F., and Zoghbi, H.Y. (2005). Math1 expression redefines the rhombic lip derivatives and reveals novel lineages within the brainstem and cerebellum. Neuron *48*, 31-43.

Watowich, S.S., and Liu, Y.J. (2010). Mechanisms regulating dendritic cell specification and development. Immunological reviews *238*, 76-92.

Weinberger, D.R. (1987). Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry *44*, 660-669.

Weinberger, D.R. (1995). From neuropathology to neurodevelopment. Lancet *346*, 552-557.

Whalen, S., Heron, D., Gaillon, T., Moldovan, O., Rossi, M., Devillard, F., Giuliano, F., Soares, G., Mathieu-Dramard, M., Afenjar, A.*, et al.* (2012). Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: Clinical score and further delineation of the TCF4 mutational spectrum. Hum Mutat *33*, 64-72.

Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N.*, et al.* (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet *382*, 1575-1586.

Wikstrom, I., Forssell, J., Penhagoncalves, M., Bergqvist, I., and Holmberg, D. (2008). A role for E2-2 at the DN3 stage of early thymopoiesis. Molecular Immunology *45*, 3302-3311.

Williams, H.J., Craddock, N., Russo, G., Hamshere, M.L., Moskvina, V., Dwyer, S., Smith, R.L., Green, E., Grozeva, D., Holmans, P.*, et al.* (2011a). Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. Hum Mol Genet *20*, 387-391.

Williams, H.J., Moskvina, V., Smith, R.L., Dwyer, S., Russo, G., Owen, M.J., and O'Donovan, M.C. (2011b). Association between TCF4 and schizophrenia does not exert its effect by common nonsynonymous variation or by influencing cis-acting regulation of mRNA expression in adult human brain. Am J Med Genet B Neuropsychiatr Genet *156B*, 781-784.

Wirgenes, K.V., Djurovic, S., Sundet, K., Agartz, I., Mattingsdal, M., Athanasiu, L., Melle, I., and Andreassen, O.A. (2010). Catechol O-methyltransferase variants and cognitive performance in schizophrenia and bipolar disorder versus controls. Schizophr Res *122*, 31-37.

Wirgenes, K.V., Sonderby, I.E., Haukvik, U.K., Mattingsdal, M., Tesli, M., Athanasiu, L., Sundet, K., Rossberg, J.I., Dale, A.M., Brown, A.A.*, et al.* (2012). TCF4 sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. Translational psychiatry *2*, e112.

Wong, E.H., So, H.C., Li, M., Wang, Q., Butler, A.W., Paul, B., Wu, H.M., Hui, T.C., Choi, S.C., So, M.T.*, et al.* (2013). Common Variants on Xq28 Conferring Risk of Schizophrenia in Han Chinese. Schizophr Bull.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., 3rd*, et al.* (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol *10*, R130.

Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet *40*, 880-885.

Xu, J., Liao, X., and Wong, C. (2010). Downregulations of B-cell lymphoma 2 and myeloid cell leukemia sequence 1 by microRNA 153 induce apoptosis in a glioblastoma cell line DBTRG-05MG. Int J Cancer *126*, 1029-1035.

Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. (2002). Allelic variation in human gene expression. Science *297*, 1143.

Yoon, S.O., and Chikaraishi, D.M. (1994). Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. J Biol Chem *269*, 18453-18462.

Yuan, P.X., Huang, L.D., Jiang, Y.M., Gutkind, J.S., Manji, H.K., and Chen, G. (2001). The mood stabilizer valproic acid activates mitogen-activated protein kinases and promotes neurite growth. J Biol Chem *276*, 31674-31683.

Yue, W.H., Wang, H.F., Sun, L.D., Tang, F.L., Liu, Z.H., Zhang, H.X., Li, W.Q., Zhang, Y.L., Zhang, Y., Ma, C.C.*, et al.* (2011). Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. Nat Genet *43*, 1228-1231.

Zaghloul, N.A., and Katsanis, N. (2010). Functional modules, mutational load and human genetic disease. Trends Genet *26*, 168-176.

Zechner, D., Fujita, Y., Hülsken, J., Müller, T., Walther, I., Taketo, M.M., Bryan Crenshaw, E., Birchmeier, W., and Birchmeier, C. (2003). β-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. Developmental Biology *258*, 406-418.

Zhang, C. (2009). MicroRNA and vascular smooth muscle cell phenotype: new therapy for atherosclerosis? Genome Med *1*, 85.

Zhang, K., Li, J.B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., Lee, J.H., Aach, J., Leproust, E.M.*, et al.* (2009). Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. Nat Methods *6*, 613-618.

Zhang, W., and Liu, H.T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res *12*, 9-18.

Zhao, Q.G., Lu, B.S., and Huang, P.T. (2005). [Functions of FANCL in primordial germ cell formation and Fanconi anemia]. Yi chuan xue bao = Acta genetica Sinica *32*, 993-1000.

Zhu, X., Gu, H., Liu, Z., Xu, Z., Chen, X., Sun, X., Zhai, J., Zhang, Q., Chen, M., Wang, K.*, et al.* (2013). Associations between TCF4 gene polymorphism and cognitive functions in schizophrenia patients and healthy controls. Neuropsychopharmacology *38*, 683-689.

Zhuang, Y., Cheng, P., and Weintraub, H. (1996). B-lymphocyte development is regulated by the combined dosage of three basic helix-loop-helix genes, E2A, E2-2, and HEB. Mol Cell Biol *16*, 2898-2905.

Zierhut, K., Bogerts, B., Schott, B., Fenker, D., Walter, M., Albrecht, D., Steiner, J., Schutze, H., Northoff, G., Duzel, E.*, et al.* (2010). The role of hippocampus dysfunction in deficient memory encoding and positive symptoms in schizophrenia. Psychiatry Res *183*, 187-194.

Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A.*, et al.* (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet *85*, 655-666.

Zweier, C., Peippo, M.M., Hoyer, J., Sousa, S., Bottani, A., Clayton-Smith, J., Reardon, W., Saraiva, J., Cabral, A., Gohring, I.*, et al.* (2007). Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet *80*, 994-1001.

APPENDIX 1

REVIEW ARTICLE

AMERICAN JOURNAL OF medical genetics PART B
Neuropsychiatric Genetics

TCF4 (*e2-2; ITF2*): A Schizophrenia-Associated Gene With Pleiotropic Effects on Human Disease

Katherinne Navarrete,[1] Inti Pedroso,[1] Simone De Jong,[2] Hreinn Stefansson,[3] Stacy Steinberg,[3] Kari Stefansson,[3,4] Roel A. Ophoff,[2,5] Leonard C. Schalkwyk,[1] and David A. Collier[1,6]*

[1]Social, Genetic and Developmental Psychiatry Centre, King's College London, Institute of Psychiatry, London, UK

[2]UCLA Center for Neurobehavioral Genetics, Gonda Center, Los Angeles, California

[3]Decode Genetics, Reykjavik, Iceland

[4]Faculty of Medicine, University of Iceland, Reykjavik, Iceland

[5]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands

[6]Eli Lilly and Company Ltd, Discovery Neuroscience Research, Erl Wood Manor, Windlesham, UK

Manuscript Received: 6 April 2012; Manuscript Accepted: 27 September 2012

Common SNPs in the transcription factor 4 (TCF4; ITF2, E2-2, SEF-2) gene, which encodes a basic Helix-Loop-Helix (bHLH) transcription factor, are associated with schizophrenia, conferring a small increase in risk. Other common SNPs in the gene are associated with the common eye disorder Fuch's corneal dystrophy, while rare, mostly de novo inactivating mutations cause Pitt-Hopkins syndrome. In this review, we present a systematic bioinformatics and literature review of the genomics, biological function and interactome of TCF4 in the context of schizophrenia. The TCF4 gene is present in all vertebrates, and although protein length varies, there is high conservation of primary sequence, including the DNA binding domain. Humans have a unique leucine-rich nuclear export signal. There are two main isoforms (A and B), as well as complex splicing generating many possible N-terminal amino acid sequences. TCF4 is highly expressed in the brain, where plays a role in neurodevelopment, interacting with class II bHLH transcription factors Math1, HASH1, and neuroD2. The Ca^{2+} sensor protein calmodulin interacts with the DNA binding domain of TCF4, inhibiting transcriptional activation. It is also the target of microRNAs, including mir137, which is implicated in schizophrenia. The schizophrenia-associated SNPs are in linkage disequilibrium with common variants within putative DNA regulatory elements, suggesting that regulation of expression may underlie association with schizophrenia. Combined gene co-expression analyses and curated protein–protein interaction data provide a network involving TCF4 and other putative schizophrenia susceptibility genes. These findings suggest new opportunities for understanding the molecular basis of schizophrenia and other mental disorders.

Key words: **Pitt-Hopkins syndrome; Fuch's corneal dystrophy; genome-wide association; protein–protein interaction; gene network**

How to Cite this Article:
Navarrete K, Pedroso I, De Jong S, Stefansson H, Steinberg S, Stefansson K, Ophoff RA, Schalkwyk LC, Collier DA. 2013. *TCF4* (*e2-2; ITF2*): A schizophrenia-associated gene with pleiotropic effects on human disease.
Am J Med Genet Part B 162B:1–16.

INTRODUCTION

The *TCF4* gene on chromosome 18q21.2 (Entrez Gene ID 6925; ensemble ENSG00000196628) encodes a basic helix-turn-helix (bHLH) Ephrussi-box (''E-box'' or E-) protein transcription factor. The gene spans 437 kb and has 41 exons. It has multiple aliases in addition to the official HUGO name *TCF4* and is also known as *E2-2*, *ITF2*, *PTHS*, *SEF2*, *SEF2-1*, *SEF2-1A*, *SEF2-1B*, *bHLHb19*, *MGC149723*, and *MGC149724*. The pleiotropic disease associations of *TCF4* include rare non-synonymous mutations and gene deletions in Pitt-Hopkins syndrome (PHS), and common risk variants for Fuch's corneal dystrophy (FCD) and schizophrenia, which are genetically distinct from each other. It should not be confused with

Additional supporting information may be found in the online version of this article.

*Correspondence to:
Prof. David A. Collier, Social, Genetic and Developmental Psychiatry Centre, King's College London, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.
E-mail: david.collier@kcl.ac.uk

Article first published online in Wiley Online Library (wileyonlinelibrary.com): 5 November 2012
DOI 10.1002/ajmg.b.32109

TCF7L2 (transcription factor 7-like 2), another transcription factor mapping to a different locus on chromosome 10q25-25.3, that plays a key role in the Wnt-signaling pathway, which has also been called *TCF4*, and has been associated with schizophrenia [Hansen et al., 2011; Alkelai et al., 2012].

In the present archival review, we performed a systematic literature search for publications describing the TCF4 protein, using the terms E2-2, ITF2, PTHS, SEF2, SEF2-1, SEF2-1A, SEF2-1B, bHLHb19, PHS, and FCD. The literature search was updated up until July 9, 2012, and references collated (see Appendix given in Supporting Information for full list of references identified for the study). Papers referring to TCF7L2 (but which used the term TCF4) were manually excluded from the list after reading the text to determine this. Several papers whose subject was TCF4 (e2-2) were incorrectly indexed with the term TCF7L2, including the original paper describing the cloning of TCF4 (e2-2). These were retained in the list. Bioinformatic analysis was performed up until April 2012.

TCF4 AND HUMAN DISEASE

Pitt-Hopkins Syndrome

Non-synonymous mutations and deletions in the TCF4 gene are a cause of PHS (MIM #610954), a rare autosomal dominant encephalopathy caused largely by de novo mutations [Whalen et al., 2012]. PHS was originally described in 1978 in two unrelated patients with mental retardation, recurrent episodes of hyperventilation, and a wide mouth [Pitt and Hopkins, 1978]. A series of additional cases helped to refine the clinical features, which are severe psychomotor delay, epilepsy, and daily bouts of diurnal hyperventilation starting in infancy; mild postnatal growth retardation; postnatal microcephaly; and distinctive facial features [Singh, 1993; Van Balkom et al., 1998; Orrico et al., 2001; Peippo et al., 2006; de Pontual et al., 2009]. Families describe their children with PHS as creative and joyful (pitthopkins.org). Clinical study finds that most have a smiling appearance, as well as anxiety, stereotypic movements, particularly of the arms, wrists, and fingers, and hyperventilation that may be triggered by emotional situations [Whalen et al., 2012]. Constipation is common, and there is co-morbidity for the developmental disorder Hirschsprung disease in PHS, a neurocristopathy characterized by the absence of enteric ganglia along a variable length of the intestine.

Using systematic genome-wide BAC-array screening, de novo microdeletions on chromosome 18q21.1, which include the TCF4 gene, were identified in PHS [Gustavsson et al., 1999; Amiel et al., 2007; Brockschmidt et al., 2007; Zweier, 2007] and fine mapping of the deleted region led to the identification of heterozygous TCF4 mutations in non-deleted cases [Amiel et al., 2007; de Pontual et al., 2009], including null alleles created by splice site, frameshift, nonsense mutations, and missense mutations which change highly conserved residues. Almost all mutations are de novo and private. Point mutations account for 40% of these, mostly causing premature stop codons, indels 30%, and de novo gene deletions 30%. Missense mutations are generally within the bHLH domain, a mutational hotspot. There may also be attenuated loss of function with a less severe phenotype: one parent carried a mosaic mutation and suffered from depression and epilepsy [de Pontual et al., 2009].

Analysis of the effects of TCF4 mutations in PHS is consistent with loss of function, with activation of a reporter construct significantly and similarly impaired for nonsense, frameshift, and missense TCF4 mutants when co-transfected with achaete-scute complex homolog 1 (HASH1; ASCL1) cDNA [de Pontual et al., 2009], a cofactor involved in activation of TCF4 DNA binding [Persson et al., 2000]. The causative mutations in TCF4 that is hemizygous deletion and inactivating mutations indicate that PHS is caused by haploinsufficiency of the gene, that is it is highly dosage-sensitive.

While TCF4 mutations are confirmed as the cause of PHS, autosomal recessive disorders resembling PHS (Pitt-Hopkins-like syndrome) are caused by mutations in NRXN1 and CNTNAP2 [Zweier et al., 2009; Blake et al., 2010]. Copy number variants (CNVs) in both of these genes are also associated with schizophrenia and other neurodevelopmental disorders [Friedman et al., 2008; Rujescu and Collier, 2009].

Schizophrenia

Genetic epidemiology studies of schizophrenia have consistently shown a strong genetic component, with estimates of heritability at up to 80% [Cardno et al., 2002; van Os and Kapur, 2009]. Multiple genetic variants have been implicated in risk, including common single nucleotide polymorphisms (SNPs) from genome-wide association with small effect-size and rare CNVs of moderate to large effects [Rujescu and Collier, 2009], as well as environmental and epigenetics factors [Pidsley and Mill, 2011]. To date, two independent genetic loci in near TCF4 have been implicated in schizophrenia, one in intron three of the gene and one between the distal end of TCF4 and the gene CCDC8, providing robust GWAS evidence for association [Stefansson et al., 2009; Ripke et al., 2011; Steinberg et al., 2011].

The initial association between TCF4 and schizophrenia was with the SNP rs9960767, in intron three of the gene ($P = 4.1 \times 10^{-9}$; OR 1.23 [1.15, 1.32]) [Stefansson et al., 2009]. A second variant (rs4309482) downstream the TCF4 gene reached a genome-wide significance level in a follow-up study ($P = 7.8 \times 10^{-9}$; 1.09 [1.06, 1.12]) [Steinberg et al., 2011], and a third variant (rs12966547) was significantly associated with schizophrenia in a GWAS mega-analysis from the psychiatric GWAS consortium ($P = 2.60 \times 10^{-10}$; OR 1.09 [1.06–1.12]) [Ripke et al., 2011]. A fourth variant, rs17512836, has also been associated with schizophrenia from the PGC schizophrenia analysis ($P = 2.35 \times 10^{-8}$); after follow-up this gave a combined P value of 1.05×10^{-6} (OR 1.23 [1.14–1.31]) [Ripke et al., 2011; broadinstitute.org/mpg/ricopili/].

The TCF4 gene, the location of GWAS associated SNPs, and their linkage disequilibrium (LD) with other SNPs at the locus is shown in Figure 1. rs12966547 is physically (1.5 kb) very close to rs4309482 and is in complete LD ($r^2 = 1$), so is effectively the same marker. rs4309482 and rs12966547 are in very weak LD with rs9960767 in the Caucasian population ($D' = 0.05$, $r^2 = 0$), that is they are showing independent association signals with schizophrenia. rs17512836 is likewise in very low LD with rs4309482 and rs12966547 ($D' = 0.02$, $r^2 = 0$), but in moderate LD with rs9960767 ($D' = 1$, $r^2 = 0.52$), indicating the association signal there is not independent.

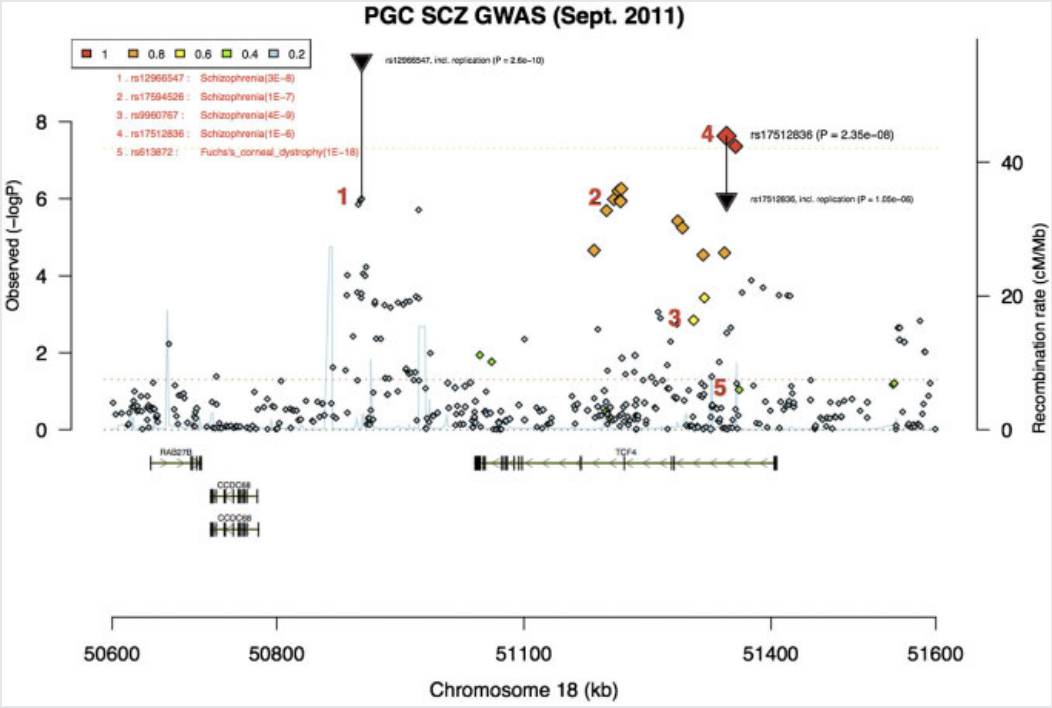


FIG. 1. Association signals between the TCF4 markers rs4309482, rs12966547, rs9960767, and rs17512836. Figure obtained from PGC schizophrenia analysis (broadinstitute.org/mpg/ricopili/).

Both rs4309482 and rs12966547 lie approximately equidistant from three genes: *CCDC68* (coiled-coil domain containing 68, aka cutaneous T-cell lymphoma associated antigen, CTLC SE57-1), a gene with unknown function; a long intergenetic non-coding RNA (ENST00000507267 or AC091103.1) overlapping TCF4 in the opposite strand; and *TCF4* itself. Therefore, in contrast with PHS mutations, it is less certain from the GWAS results which gene(s) these variants affect, as they are not obviously functional. However, a parsimonious association with TCF4 can be hypothesized given that (i) rs9960767 is in an intron of the TCF4 gene and is in low LD with the other two TCF4 associated markers, (ii) mutations in TCF4 are known to cause severe mental illness, and (iii) TCF4 is target of micro-RNA 137, which has also been associated with schizophrenia [Kwon et al., 2011; Ripke et al., 2011]. Therefore, it is reasonable that the gene mediating the genetic risk conferred by these variants is TCF4 and that the GWAS-associated variants are tagging two independent TCF4 loci which have a functional effect on the gene in some way.

However, these SNPs do not shown any association with bipolar affective disorder [Psychiatric Genetics Consortium Bipolar GWAS; Sklar et al., 2011]; http://www.broadinstitute.org/mpg/ricopili/), making them unusual in that many genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries [Williams et al., 2011a]. It is also interesting that both common and rare variants affect the same gene and confer susceptibility to brain disorders of varying severity, as it has been shown for other traits [Johansen et al., 2010; Zaghloul and Katsanis, 2010].

Two earlier genetic studies, before the advent of GWAS, examined this locus in schizophrenia: one study examined an unstable CTG repeat (CTG18.1) in an intron of TCF4 [McInnis et al., 2000] which did not show association, and a second, cytogenetic, study of families with schizophrenia or bipolar disorder found independent pericentric inversions on chromosome 18 close to the TCF4 gene (inv[18][p11.31;q21.2] and inv[18][p11.31;q21.1]) which could conceivably affect gene function [Pickard et al., 2005].

Consistent with its involvement in schizophrenia pathogenesis, blood TCF4 levels were found to be reduced during psychotic states [Kurian et al., 2011], and a small study of post-mortem cerebellar cortices of 14 schizophrenia patients and six controls also highlighted TCF4, but found up-regulation of expression in patients [Mudge et al., 2008]. However, post-mortem brain mRNA-level analyses of three brain regions, the dorsolateral prefrontal cortex (DLPFC; Brodmann area BA46) [Narayan et al., 2008], the lateral and caudal superior temporal gyrus (BA22) [Barnes et al., 2011] and the frontopolar prefrontal cortex (BA10) [Maycox et al., 2009], found no association between schizophrenia and TCF4. These discrepancies may well be explained by differences in expression patterns among tissues and cell type. Nonetheless, it remains unclear if mRNA expression changes are part of the schizophrenia pathology. In order to gather additional evidence, we examined TCF4 relative mRNA levels experimentally in lymphocytes from

cases with schizophrenia (n = 106) and controls (n = 96), and found approximately 20% lower expression in patients (fold change = 0.84 and P-value = 2.13×10^{-7}, see Supplementary Material and Methods Section for details). Williams et al. [2011b] examined TCF4 cis-regulated variation in gene expression for correlation with one of the associated SNPs, rs9960767, using an assay to detect differential allelic expression. They found no evidence for cis-regulated mRNA expression related to genotype at the schizophrenia associated SNP. However, they could not rule out temporal- or tissue-specific effects on expression. A potential limitation of these studies is that if TCF4 is highly dosage sensitive and it may be possible that small changes in mRNA and protein levels, well under those possible to detect with small samples sized studies, could increase risk of psychiatric disorders through changes in neurodevelopment. Future studies on larger sample sizes looking at specific brain regions across development will be needed to clarify the role of mRNA-level changes in schizophrenia risk.

Fuch's Corneal Dystrophy

Genetic variation within TCF4 has also been associated with FCD [Baratz et al., 2010; Li et al., 2011; Kuot et al., 2012], a common eye disease affecting about 5% of the population, and the leading cause of corneal transplant (accounting for 10% of all corneal dystrophies) [Lang and Naumann, 1987]; 80% of those affected are female. FCD is characterized by the development of collagen-free regions, which may be caused by cell death [Meek et al., 2003]. Baratz et al. [2010], in a GWAS study, showed that SNPs at the TCF4 locus were strongly associated with typical FCD, with an odds ratio of 30 for homozygotes for the disease variants. While many markers showed association, the SNPs rs17595731 and rs613872, intronic for TCF4, and rs9954153 and rs2286812, upstream of TCF4, demonstrated independent signals across the gene and gave the best case–control discrimination. rs613872 showed the strongest evidence for association ($P = 1 \times 10^{-18}$).

The FCD-associated variants appear partly independent of schizophrenia associations, even though the FCD index SNP rs613872 maps to the same intron of the gene as some of the schizophrenia associations. rs613872 shows weak to moderate LD with rs17512836 ($D' = 1$, $r^2 = 0.25$) and rs9960767 ($D' = 0.65$, 1, $r^2 = 0.20$) but negligible LD with rs4309482 ($D' = 0.07$, $r^2 = 0$) and rs12966547 ($D' = 0.07$, $r^2 = 0$). Given that schizophrenia and FCD are common in the population (about 1%) and may partially share some genetic risk, it will be worth determining whether there is any co-morbidity between these two disorders. However, there is no evidence of this from the literature.

THE TCF4 GENE

TCF4 (immunoglobulin transcription factor 2; ITF2; SEF2; OMIM *602272) was initially identified in 1990 through cloning and sequencing of the cDNA, identified as encoding a HLH transcription factor binding to Ephrussi, or E box elements (the mu-E5 motif of the immunoglobulin heavy chain enhancer and the kappa-E2 motif of the light chain enhancer) [Henthorn et al., 1990]. TCF4 was independently identified by Corneliussen et al. [1991], who characterized a protein family with homology to other bHLH transcription factors through their binding to glucocorticoid response element E-box motifs in the in the murine leukemia virus SL3-3 enhancer. Corneliussen et al. called the new protein family SEF2, for "SL3-3 enhancer factors"; TCF4 was named SEF2-1B (667 amino acids). They also identified multiple splice variants. Pscherer et al. [1996] identified the murine homologue of TCF4 (SEF2-1B; mouse Sef2) through its binding to the E box of the somatostatin receptor (SSTR2) initiator element, where it can recruit the basal transcription factor TFIIB through physical interaction with SEF2.

Comparative Genomics of TCF4

Comparative genomic analyses on 14 representative vertebrates showed that the length of the TCF4 gene greatly differs even between closely related species, for example medaka and fugu and mouse and rat, probably as a result of the extent of repeats. We observed a trend for an increase in the size of the gene from fish (36.2 kb) to human (437 kb), which could be explained under neutral evolution [Lynch, 2007]. The length of deduced human proteins from alternative TCF4 transcripts based on the Ensembl annotation is between 511 and 773 amino acids. Although the length of the deduced protein differs across species there is high sequence identity of the primary sequence, including the DNA binding domain Helix-loop-helix (Pfam domain id: PF00010) (Table I).

Supplementary Figure 1 shows an alignment of TCF4 protein sequences of 14 vertebrate species. The bHLH domain is conserved among these species as well as an element reported to play an important role the activity of the protein [Herbst and Kolligs, 2008]. Only the Ensembl database annotates a longer N-terminal region that includes a signal peptide region in a human (ENSP00000381382) and chimpanzee (ENSPTRP00000040814) isoforms (amino acids position 1–31). We performed analyses with a suite of bioinformatics tools, as described by Emanuelsson et al. [2007] to predict the function of this N-terminal region and the subcellular localization of TCF4 in human and chimpanzee. We confirmed that both the human (amino acids position 1–31) and chimpanzee (amino acids position 1–27) sequences have predicted signal peptides (Supplementary Table I; Supplementary Figs. 2–4). For the human protein, a leucine-rich nuclear export signal (amino acid position 14) was also predicted by the software NetNES [la Cour et al., 2004]. Proteins having this signal can shuttle back and forth between the nuclear and cytoplasmic compartments many times during their life cycle [Pemberton and Paschal, 2005]. In a recent study by Sepp et al. [2011] experimental evidence demonstrated TFC4 in fact does shuttle from the nucleus to the cytoplasm. They showed that full-length TCF4 isoforms that have longer N-termini were restricted to cell nucleus whereas other isoforms localized to both the nucleus and the cytoplasm, and found a nuclear localization signal in the protein in exons 8–9.

In addition to protein motifs, non-coding sequence nearby genes can have important regulatory functions, such as driving expression or modulating splicing, and represent an important source of information to understand gene function. These regulatory sequences can be identified by cross species DNA alignments. We used the VISTA genome browser [Frazer et al., 2004] and identified 365 ECRs (evolutionarily conserved sequences) between human and mouse with an average identity of approximately 75%

TABLE I. TCF4 Gene in 14 Species

		DNA		CDS		Amino acids		bHLH domain		
Species	Gene ID	Length (kb)	Repeat[a] (%)	Length (pb)	Identity[b] (%)	Length (aa)	Identity[b] (%)	Length (aa)	Identity[b] (%)	Source
Human	ENSG00000196628	413.6	23.7	2,322		773	100	55		Ensembl
Chimpanzee	ENSPTRG00000010040	412.5	22.6	2,259	99	752	96	56	91	Ensembl
Orangutan	ENSPPYG00000009175	361.4	20.3	2,016	99	671	100	55	100	Ensembl
Rhesus	ENSMMUG00000012001	360.8	19.3	2,016	98	671	99	55	100	Ensembl
Mouse	ENSMUSG00000053477	340.4	11.3	2,013	92	670	98	55	100	Ensembl
Rat	ENSRNOG00000012405	340.1	11.0	1,758	90	585	96	55	100	Ensembl
Dog	ENSCAFG00000000140	235.0	14.0	1,857	94	618	98	55	100	Ensembl
Horse	ENSECAG00000017265	345.8	17.0	2,049	96	682	98	55	100	Ensembl
Lizard	ENSACAG00000008863	268.2	3.0	1,875	85	624	94	55	100	Ensembl
Platypus	ENSOANG00000003082	157.2	23.4	1,647	87	548	96	55	100	Ensembl
Opossum	ENSMODG00000020455	223.0	59.7	1,626	90	541	97	55	100	Ensembl
Xenopus	ENSXETG00000019086	31.4	35.4	1,464	80	487	90	55	100	Ensembl
Medaka	ENSORLG00000015062	31.3	7.0	1,644	71	547	72	56	98	Ensembl
Fugu	ENSTRUG00000002569	36.2	7.0	1,581	71	526	75	56	94	Ensembl

rhesus, rhesus macaque; xenopus, xenopus tropicalis; bHLH, basic helix loop helix domain.
[a]The proportion (%) of repetitive sequences identified by the Repeat Masker in UCSC.
[b]Identity was calculated by comparing with the human TCF4 sequence.

(Supplementary Table II). When more genetically distant species (human, dog, mouse, and opossum) were employed in a cross-species sequence comparison, all exons appear highly conserved but only a few non-exonic ECRs are present (Supplementary Fig. 5). Four ECRs conserved between the human and mouse genome sequences have been tested as part of the vista enhancer project [Visel et al., 2007] (Supplementary Table III). The human element hs376 (chr18:51,240,623–51,242,077 on human genome assembly hg19) is located in intron 5 of TCF4 and was shown to drive expression of a reporter gene in the nervous system during mouse development, including the hindbrain (rhombencephalon), mid-brain (mesencephalon), and neural tube (Supplementary Fig. 6). Interestingly, we found reported genetic variation within this regulatory element (Supplementary Table IV) and several predicted transcription factor binding sites (Supplementary Fig. 7) for CREB1, ZEB1, and CUX1 among others. Interestingly, CUX1 binding to the TCF4 gene promoter has been proved experimentally by Harada et al. [2008] where chromatin affinity experiments were performed in HEK293 cell line. We also found four experimentally identified transcription factors binding sites from the ENCODE Chip-seq data: REST (NRSF), NFKB1, SPI1 (PU.1), and POU2F2 [Euskirchen et al., 2007; Robertson et al., 2007; Rozowsky et al., 2009] (Supplementary Fig. 7). Some of these transcriptions factors are known to be important in neurogenesis and neuronal plasticity, like REST and CREB1 [Dworkin and Mantamadiotis, 2010; Gao et al., 2011; Merz et al., 2011; Rossbach, 2011]. Further study of these ECRs and the potential regulatory links revealed by growing functional genomics information may shed light on the regulation of TCF4 expression and its function on the central nervous system (CNS).

THE TCF4 PROTEIN

E-proteins comprise a family of class I bHLH transcription factors homologous to the drosophila protein ''daughterless'' [Lazorchak et al., 2005; Murre, 2005] and include E12 and E47 (encoded by a single gene, E2a), HEB and TCF4 (E2-2) which form homodimers or heterodimers with other proteins in order to enable DNA binding. They have no known DNA binding activity as monomers. Class I bHLH transcription factors are expressed in a ubiquitous pattern and are able to dimerize efficiently with tissue-restricted class II factors to activate gene expression [Murre et al., 1989]. They recognize a palindromic binding site (''CANNTG''), first identified in immunoglobulin enhancers, and have a preference for C or G in the central bases. Because each partner contributes a specific DNA recognition half-site, different class I and class II heterodimers can provide distinct E box binding specificities.

Multiple alternatively spliced transcript variants that encode different TCF4 proteins have been described, with two main isoforms, E2-2A and E2-2B. E2-2B is a transcriptional repressor and a downstream target of the TCF/β-catenin complex [Petropoulos, 2000; Kolligs et al., 2002] whereas E2-2A is a transcriptional regulator with diminished or absent repressor capacity [Skerjanc et al., 1996; Furumura, 2001]. For example, the E2-2B isoform can repress the brain-specific FGF1 (fibroblast growth factor 1) by binding to an imperfect E-box in the promoter [Liu et al., 1998] while the E2-2A isoform has no activity. Likewise, interaction of E2-2B protein with MyoD can inhibit its activity by forming an inactive heterodimer [Skerjanc et al., 1996].

BIOLOGICAL FUNCTIONS AND INTERACTIONS OF TCF4

TCF4 is widely expressed in human tissues and appears to be involved in multiple biological processes, including plasmacytoid dendritic cell development [Cisse et al., 2008; Ghosh et al., 2010; Watowich and Liu, 2010], sertoli cell development [Muir et al., 2006], B and T lymphocyte development [Bergqvist et al., 2000; Engel and Murre, 2001; Quong et al., 2002; de Pooter and Kee, 2010], thymocyte development [Wikstrom et al., 2008], the epithelial-mesenchymal transition (EMT) [Sobrado et al., 2009], the maintenance of the properties of corneal epithelial stem cells during corneal epithelial development [Bian et al., 2010; Lu et al., 2012], and brain development (see below).

In EMT, a process that allows cell migration during embryonic development and tumor invasion, TCF4 can act as an E-cadherin repressor. Functional loss of the cell–cell adhesion molecule E-cadherin is an essential event in EMT and transcriptional repression has emerged as the main regulatory mechanism. E-cadherin repression mediated by TCF4 is indirect and independent of proximal E-boxes of the promoter [Sobrado et al., 2009]. Because of the potential role of TCF4 in tumor invasion, TCF4 has been examined in cancer, including estrogen receptor negative breast cancer [Appaiah et al., 2010; Brody and Witkiewicz, 2010] and is activated in human cancers with beta-catenin deficits [Kolligs et al., 2002].

TCF4 (E2-2.1) is also an essential molecular regulator of the plasmacytoid dendritic cell (PDC) lineage [Cisse et al., 2008; Nagasawa et al., 2008; Esashi and Watowich, 2009], cells which play a central role in the defense against viruses through rapidly producing interferon I [Cervantes-Barragan et al., 2012]. TCF4 shows high expression in PDCs relative to other immune cells and regulates terminal development and interferon production by controlling PDC-related genes. Impairment of TCF4 (E2-1.1) leads to a reduced number and abnormal cell surface of PDCs, which is thought to be the basis for persistent respiratory infection in patients with PHS.

TCF4 may also involved in regulation of calcium signaling in cells through its interaction with calmodulin. TCF4 along with the E2A proteins E12 and E47 bind calmodulin in vitro at physiologically relevant calmodulin concentrations [Corneliussen et al., 1994] in an interaction modulated by sequences directly N-terminal to the basic domain. This results in inhibition of DNA binding in vitro. Increased intracellular Ca^{2+} concentration potently inhibits the transcriptional activity of TCF4 on an E-box containing reporter plasmid, as does calmodulin overexpression. Thus Ca^{2+} signaling may inhibit the transcriptional activity of E-proteins, including TCF4, by establishing an E-protein-calmodulin complex that prevents the E-protein from interaction with its target DNA [Saarikettu, 2004].

TCF4 expression in the brain is likely to play a major role in neurodevelopment, especially as haploinsufficiency has been associated with PHS and common variants in or near the gene confer risk of schizophrenia. However, there is a paucity of information on the role of TCF4 in the CNS. TCF4 is known to be important in the differentiation of glial cells, especially the maturation of oligodendrocyte progenitors In the postnatal spinal cord TCF4 expression is exclusive to oligodendrocyte lineage cells and decreases as myelination proceeds, whereas in the brain, TCF4 is also expressed in neurons, especially in the thalamus [Fu et al., 2009].

TCF4 is known to interact with several class II bHLH transcription factor genes involved in neurodevelopment: Math1, a proneural protein expressed in the differentiating neuroepithelium [Ross et al., 2003; Flora et al., 2007; Gohlke et al., 2008]; HASH1, a protein necessary for the formation of distinct neuronal circuits within the CNS, especially the telencephalon [Persson et al., 2000]; neuroD2, which plays important roles in neuronal differentiation and survival [Ravanpay and Olson, 2008; and Id1, a homologue of the drosophila daughterless and emc loci, which are required for correct patterning in neurogenesis [Einarson and Chao, 1995]. TCF4 is also the only transcription factor showing complete co-expression with Olig2 [Fu et al., 2009], an essential regulator of ventral neuroectodermal progenitor cell fate and a requirement for oligodendrocyte and motor neuron development [Othman et al., 2011; Panman et al., 2011]. Taken together, these data suggest that TCF4 is involved in an intricate combinatorial regulatory circuit during CNS development. Ravasi et al. [2010] support a model in which the combinatorial nature of transcription factors rather than their expression in isolation is the main determinant of tissue specificity and the functionality of transcription factors complexes is associated with their conservation across species. Ravasi et al. [Ravasi et al., 2010] show that TCF4 is part of conserved transcription factor complex of proteins co-expressed in cerebellum across human and mouse.

Math1 is highly expressed in proliferating progenitors of the rhombic lip, a specialized neuroepithelium in the developing dorsal hindbrain [Akazawa et al., 1995; Ben-Arie et al., 1996], and is also required for the differentiation of most cerebellar and pre-cerebellar structures [Ben-Arie et al., 1997; Wang et al., 2005]—for example, Math1 knockout mice lack all cerebellar granule neurons. It is important to note that cerebellar expression of TCF4 is very high [Sepp et al., 2011] and could relate to Math1. TCF4 also forms a functional complex with HASH-1, and thus appears to be involved in the development of specific parts of the central and peripheral nervous systems [Persson et al., 2000]. The HASH-1/TCF4 complex binds an E-box (CACCTG) in vitro and is one of the major HASH-1 interacting proteins in extracts from neuroblastoma cells. HASH-1 (MASH-1 in mouse) is a transcription factor homologous to the drosophila "achaete-scute homolog-1" and is essential for correct development of olfactory and most peripheral autonomic neurons, and for the formation of distinct neuronal circuits within the CNS [Persson et al., 2000].

The E proteins E12, HEB, and TCF4 all interact with neuroD2, and heterodimers bind a neuroD2 preferred E-box. These complexes have a comparable ability to induce neurogenesis, indicating a similar baseline biological activity [Ravanpay and Olson, 2008]. NeuroD2 has been reported to play a critical role in the induction of neuronal differentiation, the promotion of neuronal survival, and development of thalamocortical communication [Farah et al., 2000; Olson et al., 2001; Ince-Dunn et al., 2006; Lin et al., 2006; Noda et al., 2006].

Brockschmidt et al. [2011] performed knockdown of TCF4 in zebrafish embryos, which resulted in developmental delay or defects

in terminal differentiation of brain and eyes. The authors conclude that TCF4-knockdown in zebrafish embryos did not seem to affect early neural patterning and regionalization of the forebrain, but may be involved in later aspects of neurogenesis and differentiation.

HUMAN TCF4 EXPRESSION

Inspection of public databases showed that TCF4 is expressed at detectable levels across many tissues (Supplementary Figs. 8 and 9). However, it expression is high in prefrontal cortex, fetal brain, and BDCA4+ dendritic cells. We explored the Human Brain Transcriptome dataset, which has expression data of TCF4 across several brain regions and developmental stages [Kang et al., 2011], and we also examined the Sestan Lab Human Brain Atlas Microarrays dataset [Johnson et al., 2009] to explore the expression pattern of TCF4 in human brain. We found that TCF4 has higher expression in neocortex and hippocampus compared with striatum, thalamus, and cerebellum and that its expression seems to be higher at early stages of development (Fig. 2; Supplementary Table V). These brain areas have been implicated with schizophrenia by neuropsychological neuroimaging and post mortem histology and neurochemical studies [Heckers, 1997; Harrison, 2004; Boyer et al., 2007; Casanova et al., 2008; Lewis and Gonzalez-Burgos, 2008; Zierhut et al., 2010].

de Pontual et al. [2009] found that TCF4 is highly expressed throughout the human CNS development and the sclerotomal component of the somites from Carnegie stage C13 28–32 days postfertilization (dpf). TCF4 expression continues into adult life, expressed in lymphocytes, fibroblasts, gut, muscle, but not the heart. Sepp et al. [2011] used RT-PCR to detect TCF4 transcripts and also found widespread expression, including testis, placenta, prostate, and in postmortem adult human brain, where they found particularly high expression, especially in the cerebellum.

TCF4 Transcripts

The TCF4 gene has numerous exons and produces several transcripts. Currently two human transcripts (isoforms TCF4A [ITF2A; SEF2-1A] and the canonical sequence TCF4B [ITF2B; SEF2-1B]; accession numbers NM_001083962.1 and NM_003199.2) are commonly referred to in the literature, along with a third known isoform SEF2-1D (TCF4D). Likewise, two transcripts of human TCF4 are annotated in the NCBI Entrez database, but four in Ensembl, 10 in UCSC and 38 in NCBI AceView database (version April 2007) [Thierry-Mieg and Thierry-Mieg, 2006]. More than 459 mRNA or EST sequences are related to human TCF4. Forty one splice variants are recorded at Ensembl's latest version (version 66), many of which are derived from RNA-seq (de novo sequencing of transcripts [Ozsolak and Milos, 2011]) experiments from the Illumina Body Map and ENCODE projects and therefore the splice junctions are likely to be real. Increasing availability of RNA-seq data will help to elucidate the full coding potential of the gene as well as possible transcriptional biases for different isoforms on specific tissues.

Transcripts have been experimentally verified in mice [Brzozka et al., 2010], rat [Yoon and Chikaraishi, 1994], and humans [Pscherer et al., 1996]. Sepp et al. [2011] used RT-PCR in a detailed bioinformatic and experimental study of TCF4. They demonstrated the existence of several 5′ exons that potentially yield 18 different N-termini in TCF4 proteins, and alternative splicing of several

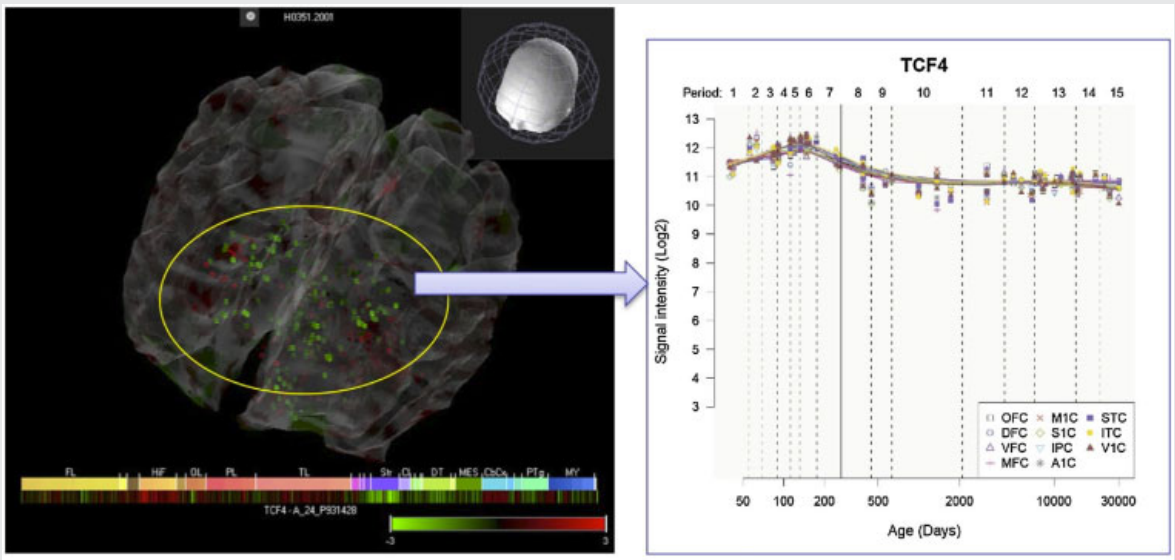


FIG. 2. Expression of TCF4 across several brain regions and developmental stages. Data obtained from the Human Brain Transcriptome dataset and the Sestan Lab Human Brain Atlas Microarrays dataset. TCF4 has higher expression in neocortex and hippocampus compared with striatum, thalamus, and cerebellum and it expression seems to be higher at early stages of development.

internal exons that may give rise to further isoforms. All the isoforms contain transactivation domain AD2 and the bHLH domain, however, only the isoforms with the longest N-terminal have the amino-terminal transactivation domain AD1. AD1 and AD2 are able to activate transcription individually and together they act synergistically [Sepp et al., 2011].

Some of these isoforms contain a nuclear localization signal (NLS) and are only found in the nucleus, whereas others are also found in the cytoplasm. Isoforms without an NLS may rely on binding to bHLH partner proteins, such as NeuroD which contains an NLS, in order to enter the nucleus. Sepp et al. [2011] also conclude that isoforms lacking NLS can be exported from the nucleus by heterodimerization with nuclear export signal (NES) containing partner proteins. In the present study, we also detected a leucine-rich NES (amino acid position 14), which could also explain its export from the nucleus. In Supplementary Table VI, we provide identifier mapping for isoforms across different databases that can be useful for researchers aiming to use information across these databases.

Potential miRNA Regulation

TCF4 is known to be regulated by at least two differentially expressed microRNAs, miR-221 and miR-222, which regulate plasmacytoid versus conventional dendritic cell development [Kuipers et al., 2010]. MicroRNAs are likely to confer tissue-specific fine-tuning of TCF4 expression, for example in neurons. In order to explore this, we used the TargetSCAN [Lewis et al., 2005] database to identify predicted miRNA binding sites in the TCF4 UTR conserved regions. There were 45 miRNA predicted binding sites conserved across vertebrates. Of them, 37 have been associated with human diseases and 30 specifically with CNS disorders (Table II). Interestingly, hsa-mir-106a/b, hsa-mir-7-1/2/3, hsa-mir-92a-1/2, hsa-mir-92b, and hsa-mir-20a/b have been associated with schizophrenia and autism [Perkins et al., 2007; Abu-Elneel et al., 2008; Talebizadeh et al., 2008]. Furthermore, a genetic variant near hsa-mir-137 has been associated with schizophrenia at genome-wide significant level [Ripke et al., 2011]. The targeting of has-mir-137 with TCF4 has been confirmed by Kwon et al. [2011]. hsa-mir-155 has been also validated experimentally according to the database mirWalk [Dweep et al., 2011] (http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/mirnatargetpub.html).

SNPs in TCF4

We found 18,636 SNPs (db130) in the locus, which we defined as the gene boundaries plus approximately 2 Mb in each direction to cover all proximal haplotype blocks (chr 18: 49.019.029–52.414.148, human assembly hg18). Of them, 4,417 and 6,762 have been genotyped by the HapMap and 1,000 genomes project, respectively, and 2,706 in both. We used Galaxy tools [Blankenberg et al., 2010] to intersect these SNPs with a set of putative and validated DNA functional elements and found that 12,500 SNPs (67%) are within potentially regulatory regions, including multispecies conserved sequences and known protein binding sites, and 64 were within exons (Fig. 3). Among the latter, 29 were validated non-synonymous SNPs, as listed in the UCSC genome browser (Supplementary Table IX).

Williams et al. [2011b] used data from the 1,000 genomes project to identify non-synonymous coding variants in the TCF4 gene, but did not find any evidence for association and conclude that association between schizophrenia and TCF4 is not mediated by a relatively common non-synonymous variant. They also performed allele-specific expression analysis using human post mortem brain samples, but found no evidence for cis-regulated mRNA expression related to genotype at the schizophrenia-associated SNP.

Eighty-eight out of the 4,417 SNPs genotyped by HapMap were both in high LD ($r^2 > 0.8$) with the SNPs associated with schizophrenia, rs4309482 and rs9960767, and within potentially regulatory sites (Table III), suggesting that the disease-susceptibility alleles may capture variants influencing gene expression. To test this we looked at eQTL dataset at http://eqtl.uchicago.edu/Home.html [Veyrieras et al., 2008] where we found one eQTL mapping TCF4 gene region, rs1261085. However, this eQTL showed low LD with the three schizophrenia-associated SNPs, for example with rs996076 had an r^2 of 0.073, and an r^2 of 0.003 with rs4309482 and rs12966547.

CO-EXPRESSION AND PROTEIN–PROTEIN INTERACTION NETWORKS OF TCF4

We used network biology information to help understand the function of TCF4 by using higher-level biological information [Geschwind and Konopka, 2009]. Gene co-expression networks derived from gene-expression analysis of human brains [Oldham et al., 2008] revealed 139 genes tightly co-expressed with TCF4. These genes were analyzed by gene-set enrichment analysis with MetaCore (GeneGO, Inc.) showing that they are characterized by specific biological processes including Alpha-2 adrenergic receptor activation of mitogen-activated protein kinase 1 and 3 (MAPK1/3) (P-value $= 1.5 \times 10^{-3}$), WNT signaling pathway (P-value 5.9×10^{-3}) and cell adhesion/synaptic contact (P-value 3.0×10^{-3}) (Supplementary Table VII). It is interesting that some of them have been previously associated with schizophrenia, such as the WNT signaling pathway [Cotter et al., 1998; Miyaoka et al., 1999; Freyberg et al., 2010] and cell adhesion/synaptic contact [Vawter, 2000; Gross et al., 2003; Eastwood, 2004; Stephan et al., 2009; Hayashi-Takagi and Sawa, 2010; O'Dushlaine et al., 2011].

In addition, we used GeneGO Metacore (GeneGo, Inc.) manually curated information to identify TCF4 interacting partners, based on physical or regulatory interactions, that are also known to be associated with schizophrenia. We identified six proteins, namely TP53, NR3C1, AR, CTNNB1, ESR1, and EGFR, known to regulate or interact with TCF4 and further seven from among the co-expression partners, that is GSK3β, NRXN1, GABRA1, MILL, GRIA3, GNAL, and ADRA2A. We used the manually curated interactions and the co-expression results to build a network of schizophrenia susceptibility genes around TCF4 (Fig. 4). This sheds light on its relationships with known disease susceptibility genes.

TABLE II. Predicted Conserved miRNA Target Sites in TCF4

miRNA	Disease	Reference
hsa-mir-106a/hsa-mir-106b	Schizophrenia, alzheimer disease, autistic disorder, glioma	[Abu-Elneel et al., 2008; Hebert et al., 2009; Lavon et al., 2010; Perkins et al., 2007]
hsa-mir-7-1/hsa-mir-7-2/hsa-mir-7-3	Schizophrenia, autistic disorder	[Abu-Elneel et al., 2008; Perkins et al., 2007]
hsa-mir-92a-1/hsa-mir-92a-2/hsa-mir-92b	Schizophrenia, autistic disorder, medulloblastoma, brain neoplasms	[Nass et al., 2009; Northcott et al., 2009; Perkins et al., 2007; Talebizadeh et al., 2008; Uziel et al., 2009]
hsa-mir-20a/hsa-mir-20b	Schizophrenia, alzheimer disease, medulloblastoma, glioma, glioblastoma	[Ernst et al., 2010; Hebert et al., 2009; Lavon et al., 2010; Northcott et al., 2009; Perkins et al., 2007; Uziel et al., 2009]
hsa-mir-93	Autistic disorder, glioma	[Abu-Elneel et al., 2008; Lavon et al., 2010]
hsa-mir-129-1/hsa-mir-129-2	Autistic disorder	[Abu-Elneel et al., 2008]
hsa-mir-363	Autistic disorder	[Talebizadeh et al., 2008]
hsa-mir-193b	Autistic disorder	[Abu-Elneel et al., 2008]
hsa-mir-17	Alzheimer disease, medulloblastoma, glioma, glioblastoma	[Ernst et al., 2010; Hebert et al., 2009; Lavon et al., 2010; Northcott et al., 2009; Uziel et al., 2009]
hsa-mir-139	Neurodegenerative diseases	[Saba et al., 2008]
hsa-mir-155	Down syndrome, cerebral hemorrhage, stroke, glioblastoma	[Care et al., 2007; Li et al., 2009; Liu et al., 2010; Sethupathy et al., 2007]
hsa-mir-124-1/hsa-mir-124-2/hsa-mir-124-3	Glioblastoma, medulloblastoma, stroke, cocaine-related disorders	[Chandrasekar and Dreyer, 2009; Laterza et al., 2009; Pierson et al., 2008; Silber et al., 2008]
hsa-mir-128-1/hsa-mir-128-2	Neuroblastoma	[Evangelisti et al., 2009]
hsa-mir-137	Glioblastoma	[Silber et al., 2008]
hsa-mir-141	ACTH-secreting pituitary adenoma	[Amaral et al., 2009]
hsa-mir-145	ACTH-secreting pituitary adenoma, atherosclerosis	[Amaral et al., 2009; Zhang, 2009]
hsa-mir-183	Glioma	[Lavon et al., 2010]
hsa-mir-200a	Meningioma	[Poethig, 2009]
hsa-mir-221	Astrocytoma	[Conti et al., 2009]
hsa-mir-367	Glioma	[Lavon et al., 2010]

miRNA potentially regulating TCF4 and associated with CNS diseases.

SUMMARY AND FUTURE DIRECTIONS

TCF4 is still a poorly characterized gene in terms of function and biology. However, the compelling evidence for genetic association of TCF4 with schizophrenia calls for additional detailed studies. For example, the current literature only refers to three protein isoforms at the most, but there is experimental evidence for many more [Sepp et al., 2011]. Future studies will benefit from a careful dissection of the functional differences between TCF4 splice variants, especially since they differ in functional motifs. This potential array of different protein products from the TCF4 gene may account for its pleiotropic associations with disease.

Although the role of TCF4 during development has not been elucidated, it seems plausible, based on its known interactions with

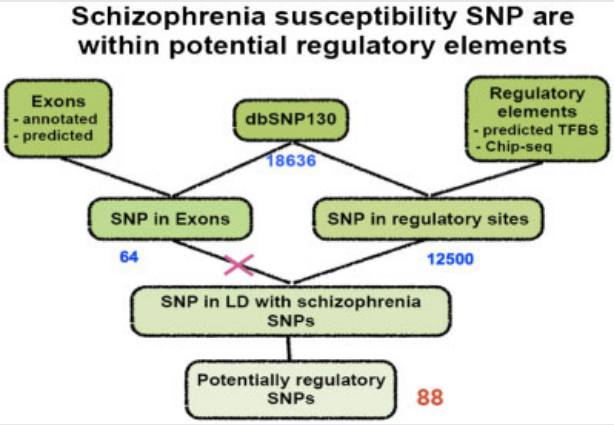


FIG. 3. Summary of the strategy used to find the 88 potentially regulatory SNPs in LD with the SNPs associated with schizophrenia.

Math1, HASH1, and neuroD2, to suggest that TCF4 plays an important role during CNS development. Because of its interaction with calmodulin and involvement in calcium signaling, TCF4 may also be involved in acute functional effects in neurons. The advent of second generation sequencing technologies will allow a more thorough study of transcription factor binding site targets of TCF4, using methods such as ChIP-seq, as well as isoform abundance at different points during development of CNS tissues using RNA-Seq. These analyses will reveal further information on the TCF4 biological regulatory network, the role of the protein in brain development and function and, in turn, TCF4's involvement in psychiatric and neurodevelopmental disorders.

The role of TCF4 in brain neural function, especially in relation to schizophrenia, also needs to be elucidated. For example, preliminary studies are examining the role of TCF4 polymorphisms on putative endophenotypes of schizophrenia have shown an effect on sensory gating, as measured by the P50 suppression of the auditory evoked potential [Quednow et al., 2012], and on prepulse inhibition of the acoustic startle response, a different measurement of sensorimotor gating [Quednow et al., 2011]. These studies may provide important information on the relationship between TCF4 risk polymorphisms and neuropsychiatric dysfunctions in schizophrenia.

Our scan for putative functional genetic variants revealed that several common SNPs are both within likely functional DNA elements and in LD with SNPs associated to schizophrenia. In line with these findings, one of these elements is known to drive gene expression in mouse CNS. Additional studies exploring the function of these DNA sequences can provide clues as to which transcription factors are regulating TCF4 and in turn how TCF4 is being integrated into wider regulatory processes. In addition, TCF4 transcripts have several potential binding sites for miRNAs whose expression has been associated with mental illnesses. Additional validation of these regulatory links is warranted and could provide important clues between TCF4 and other genes underlying risk of mental disorders. Thus, the information we have compiled suggests it may be possible to use additional targeted functional studies to start drawing a functional network relating TCF4 to other key molecular players in brain development and mental illness. Furthermore, we showed that by compiling gene-expression data and protein–protein interaction information TCF4 is readily linked to other genes associated with schizophrenia. A major challenge will be to understand the mechanisms relating these genes and in which context, for example cell type or developmental stage, their interactions are important as mediators of risk for schizophrenia. This effort will need to be undertaken by a wide community applying a plethora of experimental approaches. Nonetheless, this represents that the start of a path lead to biological understanding and possibly new treatments for schizophrenia and other psychiatric disorders.

TABLE III. Putative Functional SNPs in LD With Schizophrenia Associated SNPs

SNP 1	SNP ID	Position	r-square	Functional element
rs4309482	rs4309482	50901467	1	Nucleosome occupancy (MEC)
	rs12966547	50903015	1	phastConsElements28wayPlacMammal, phastConsElements28way, Nucleosome occupancy (MEC)
	rs4801131	50903698	1	Nucleosome occupancy (MEC)
	rs4131791	50898869	1	Nucleosome occupancy (MEC)
rs9960767	rs9960767	51306000	1	Nucleosome occupancy (MEC), Sunny RBP
	rs12327270	51319071	1	phastConsElements28wayPlacMammal, phastConsElements28way, Nucleosome occupancy (MEC), Regulatory Potential 7x, Gis Chip (c-Myc), Sunny RBP
	rs10401120	51343496	0.867	Nucleosome occupancy (MEC), Sunny RBP

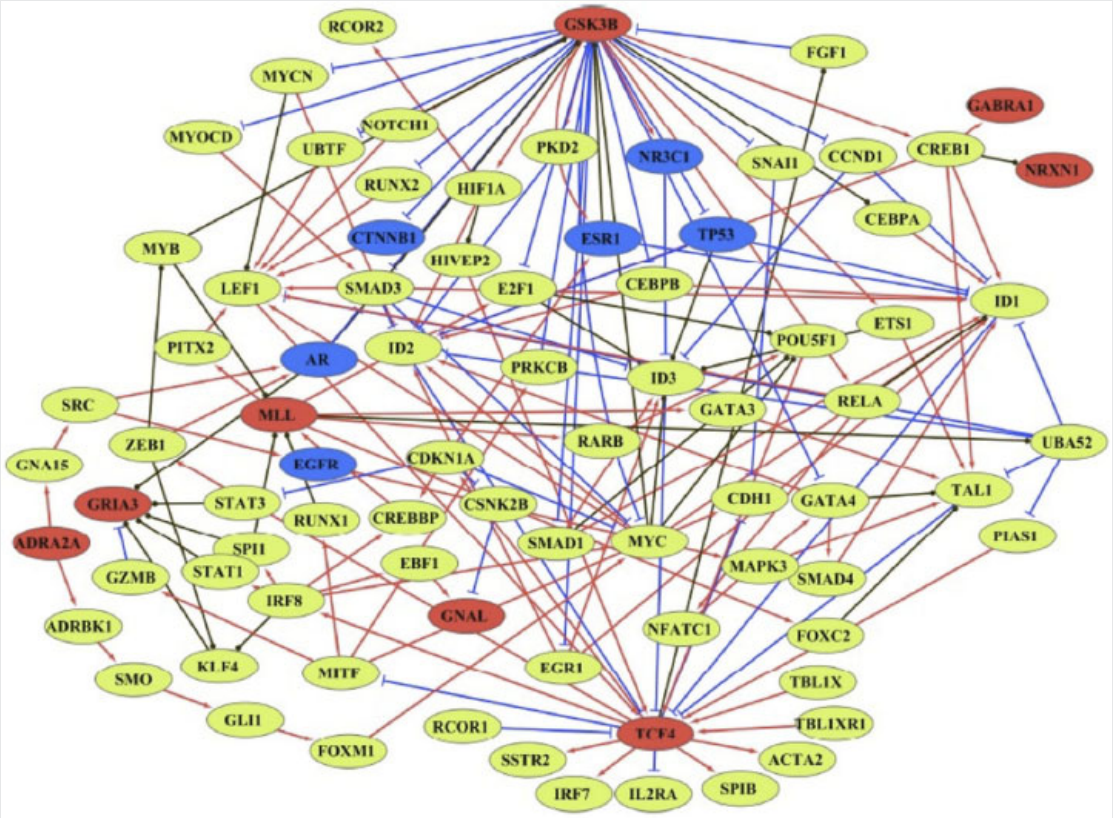


FIG. 4. Network relationships between TCF4, genes that are coexpressed with TCF4 in human cortex and genes that are associated with schizophrenia. Red is the genes coexpressed with TCF4 and associated with schizophrenia and blue is the genes associated with schizophrenia. The arrows show the direction of the interaction and the color if it is activation (green), inhibition (red), and unspecified (black).

REFERENCES

Abu-Elneel K, Liu T, Gazzaniga FS, Nishimura Y, Wall DP, Geschwind DH, Lao K, Kosik KS. 2008. Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics 9:153–161.

Akazawa C, Ishibashi M, Shimizu C, Nakanishi S, Kageyama R. 1995. A mammalian helix-loop-helix factor structurally related to the product of drosophila proneural gene atonal is a positive transcriptional regulator expressed in the developing nervous system. J Biol Chem 270:8730–8738.

Alkelai A, Greenbaum L, Lupoli S, Kohn Y, Sarner-Kanyas K, Ben-Asher E, Lancet D, Macciardi F, Lerer B. 2012. Association of the type 2 diabetes mellitus susceptibility gene, TCF7L2, with schizophrenia in an Arab-Israeli family sample. PLoS ONE 7(1):e29228.

Amaral FC, Torres N, Saggioro F, Neder L, Machado HR, Silva WA Jr, Moreira AC, Castro M. 2009. MicroRNAs differentially expressed in ACTH-secreting pituitary tumors. J Clin Endocrinol Metab 94:320–323.

Amiel J, Rio M, de Pontual L, Redon R, Malan V, Boddaert N, Plouin P, Carter NP, Lyonnet S, Munnich A, et al. 2007. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet 80: 988–993.

Appaiah H, Bhat-Nakshatri P, Mehta R, Thorat M, Badve S, Nakshatri H. 2010. ITF2 is a target of CXCR4 in MDA-MB-231 breast cancer cells and is associated with reduced survival in estrogen receptor-negative breast cancer. Cancer Biol Ther 10:600–614.

Baratz KH, Tosakulwong N, Ryu E, Brown WL, Branham K, Chen W, Tran KD, Schmid-Kubista KE, Heckenlively JR, Swaroop A, et al. 2010. E2-2 protein and Fuchs's corneal dystrophy. N Engl J Med 363:1016–1024.

Barnes MR, Huxley-Jones J, Maycox PR, Lennon M, Thornber A, Kelly F, Bates S, Taylor A, Reid J, Jones N, et al. 2011. Transcription and pathway analysis of the superior temporal cortex and anterior prefrontal cortex in schizophrenia. J Neurosci Res 89:1218–1227.

Ben-Arie N, McCall AE, Berkman S, Eichele G, Bellen HJ, Zoghbi HY. 1996. Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis. Hum Mol Genet 5:1207–1216.

Ben-Arie N, Bellen HJ, Armstrong DL, McCall AE, Gordadze PR, Guo Q, Matzuk MM, Zoghbi HY. 1997. Math1 is essential for genesis of cerebellar granule neurons. Nature 390:169–172.

Bergqvist I, Eriksson M, Saarikettu J, Eriksson B, Corneliussen B, Grundstrom T, Holmberg D. 2000. The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. Eur J Immunol 30:2857–2863.

Blake DJ, Forrest M, Chapman RM, Tinsley CL, O'Donovan MC, Owen MJ. 2010. TCF4, schizophrenia, and Pitt-Hopkins syndrome. Schizophr Bull 36:443–447.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010. Galaxy: A web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol 10:1–21.

Boyer P, Phillips JL, Rousseau FL, Ilivitsky S. 2007. Hippocampal abnormalities and memory deficits: New evidence of a strong pathophysiological link in schizophrenia. Brain Res Rev 54:92–112.

Brockschmidt A, Todt U, Ryu S, Hoischen A, Landwehr C, Birnbaum S, Frenck W, Radlwimmer B, Lichter P, Engels H, et al. 2007. Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. Hum Mol Genet 16:1488–1494.

Brockschmidt A, Filippi A, Issa P, Charbel P, Nelles M, Urbach H, Eter N, Driever W, Weber RG. 2011. Neurologic and ocular phenotype in Pitt-Hopkins syndrome and a zebrafish model. Hum Genet 130:645–655.

Brody JR, Witkiewicz AK. 2010. CXCR4 signaling identifies a role for IFT2 in ER-negative breast cancers. Cancer Biol Ther 10(6):615–616.

Brzozka MM, Radyushkin K, Wichert SP, Ehrenreich H, Rossner MJ. 2010. Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. Biol Psychiatry 68:33–40.

Cardno AG, Rijsdijk FV, Sham PC, Murray RM, McGuffin P. 2002. A twin study of genetic relationships between psychotic symptoms. Am J Psychiatry 159:539–545.

Care A, Catalucci D, Felicetti F, Bonci D, Addario A, Gallo P, Bang ML, Segnalini P, Gu Y, Dalton ND, et al. 2007. MicroRNA-133 controls cardiac hypertrophy. Nat Med 13:613–618.

Casanova MF, Kreczmanski P, Trippe J II, Switala A, Heinsen H, Steinbusch HW, Schmitz C. 2008. Neuronal distribution in the neocortex of schizophrenic patients. Psychiatry Res 158:267–277.

Cervantes-Barragan L, Lewis KL, Firner S, Thiel V, Hugues S, Reith W, Ludewig B, Reizis B. 2012. Plasmacytoid dendritic cells control T-cell response to chronic viral infection. Proc Natl Acad Sci USA 109: 3012–3017.

Chandrasekar V, Dreyer JL. 2009. microRNAs miR-124, let-7d and miR-181a regulate cocaine-induced plasticity. Mol Cell Neurosci 42:350–362.

Cisse B, Caton M, Lehner M, Maeda T, Scheu S, Locksley R, Holmberg D, Zweier C, Denhollander N, Kant S. 2008. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell 135:37–48.

Conti A, Aguennouz M, La Torre D, Tomasello C, Cardali S, Angileri FF, Maio F, Cama A, Germano A, Vita G, et al. 2009. miR-21 and 221 upregulation and miR-181b downregulation in human grade II-IV astrocytic tumors. J Neurooncol 93:325–332.

Corneliussen B, Thornell A, Hallberg B, Grundstrom T. 1991. Helix-loop-helix transcriptional activators bind to a sequence in glucocorticoid response elements of retrovirus enhancers. J Virol 65:6084–6093.

Corneliussen B, Holm M, Waltersson Y, Onions J, Hallberg B, Thornell A, Grundstrom T. 1994. Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. Nature 368:760–764.

Cotter D, Kerwin R, Al-Sarraji S, Brion JP, Chadwich A, Lovestone S, Anderton B, Everall I. 1998. Abnormalities of Wnt signalling in schizophrenia-evidence for neurodevelopmental abnormality. Neuroreport 9:1379–1383.

de Pontual LC, Mathieu Y, Golzio C, Rio MN, Malan VR, Boddaert N, Soufflet C, Picard C, Durandy A, Dobbie A, et al. 2009. Mutational, functional, and expression studies of theTCF4 gene in Pitt-Hopkins syndrome. Hum Mutat 30:669–676.

de Pooter RF, Kee BL. 2010. E proteins and the regulation of early lymphocyte development. Immunol Rev 238:93–109.

Dweep H, Sticht C, Pandey P, Gretz N. 2011. miRWalk-database: Prediction of possible miRNA binding sites by ''walking'' the genes of three genomes. J Biomed Inform 44:839–847.

Dworkin S, Mantamadiotis T. 2010. Targeting CREB signalling in neurogenesis. Expert Opin Ther Targets 14:869–879.

Eastwood SL. 2004. The synaptic pathology of schizophrenia: Is aberrant neurodevelopment and plasticity to blame? Int Rev Neurobiol 59:47–72.

Einarson MB, Chao MV. 1995. Regulation of Id1 and its association with basic helix-loop-helix proteins during nerve growth factor-induced differentiation of PC12 cells. Mol Cell Biol 15:4175–4183.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953–971.

Engel I, Murre C. 2001. The function of E- and Id proteins in lymphocyte development. Nat Rev Immunol 1:193–199.

Ernst A, Campos B, Meier J, Devens F, Liesenberg F, Wolter M, Reifenberger G, Herold-Mende C, Lichter P, Radlwimmer B. 2010. De-repression of CTGF via the miR-17-92 cluster upon differentiation of human glioblastoma spheroid cultures. Oncogene 29:3411–3422.

Euskirchen GM, Rozowsky JS, Wei CL, Lee WH, Zhang ZD, Hartman S, Emanuelsson O, Stolc V, Weissman S, Gerstein MB, et al. 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: Comparison of array- and sequencing-based technologies. Genome Res 17:898–909.

Evangelisti C, Florian MC, Massimi I, Dominici C, Giannini G, Galardi S, Bue MC, Massalini S, McDowell HP, Messi E, et al. 2009. MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness. FASEB J 23:4276–4287.

Farah MH, Olson JM, Sucic HB, Hume RI, Tapscott SJ, Turner DL. 2000. Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. Development 127:693–702.

Flora A, Garcia JJ, Thaller C, Zoghbi HY. 2007. The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. Proc Natl Acad Sci 104:15382–15387.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: Computational tools for comparative genomics. Nucleic Acids Res 32:W273–W279.

Freyberg Z, Ferrando SJ, Javitch JA. 2010. Roles of the Akt/GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. Am J Psychiatry 167:388–396.

Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, Knoers NV, Cahn W, Kahn RS, Edelmann L, et al. 2008. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry 13:261–266.

Fu H, Cai J, Clevers H, Fast E, Gray S, Greenberg R, Jain MK, Ma Q, Qiu M, Rowitch DH, Taylor CM, Stiles CD. 2009. A genome-wide screen for spatially restricted expression patterns identifies transcription factors that regulate glial development. J Neurosci 29:11399–11408.

Furumura M. 2001. Involvement of ITF2 in the transcriptional regulation of melanogenic genes. J Biol Chem 276:28147–28154.

Gao Z, Ure K, Ding P, Nashaat M, Yuan L, Ma J, Hammer RE, Hsieh J. 2011. The master negative regulator REST/NRSF controls adult neurogenesis by restraining the neurogenic program in quiescent stem cells. J Neurosci 31:9772–9786.

Geschwind DH, Konopka G. 2009. Neuroscience in the era of functional genomics and systems biology. Nature 461:908–915.

Ghosh HS, Cisse B, Bunin A, Lewis KL, Reizis B. 2010. Continuous expression of the transcription factor e2-2 maintains the cell fate of mature plasmacytoid dendritic cells. Immunity 33:905–916.

Gohlke JM, Armant O, Parham FM, Smith MV, Zimmer C, Castro DS, Nguyen L, Parker JS, Gradwohl G, Portier CJ, et al. 2008. Characterization of the proneural gene regulatory network during mouse telencephalon development. BMC Biol 6:15.

Gross J, Grimm O, Meyer J, Lesch KP. 2003. The cadherin hypothesis of schizophrenia. Fortschr Neurol Psychiatr 71:84–88.

Gustavsson P, Kimber E, Wahlstrom J, Anneren G. 1999. Monosomy 18q syndrome and atypical Rett syndrome in a girl with an interstitial deletion (18)(q21.1q22.3). Am J Med Genet 82:348–351.

Hansen T, Ingason A, Djurovic S, Melle I, Fenger M, Gustafsson O, Jakobsen KD, Rasmussen HB, Tosato S, Rietschel M, et al. 2011. At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. Biol Psychiatry 70:59–63.

Harada R, Vadnais C, Sansregret L, Leduy L, Berube G, Robert F, Nepveu A. 2008. Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes. Nucleic Acids Res 36:189–202.

Harrison PJ. 2004. The hippocampus in schizophrenia: A review of the neuropathological evidence and its pathophysiological implications. Psychopharmacology (Berl) 174:151–162.

Hayashi-Takagi A, Sawa A. 2010. Disturbed synaptic connectivity in schizophrenia: Convergence of genetic risk factors during neurodevelopment. Brain Res Bull 83:140–146.

Hebert SS, Horre K, Nicolai L, Bergmans B, Papadopoulou AS, Delacourte A, De Strooper B. 2009. MicroRNA regulation of Alzheimer's amyloid precursor protein expression. Neurobiol Dis 33:422–428.

Heckers S. 1997. Neuropathology of schizophrenia: Cortex, thalamus, basal ganglia, and neurotransmitter-specific projection systems. Schizophr Bull 23:403–421.

Henthorn P, Carrick-Walmsley R, Kadesch T. 1990. Sequence of the cDNA encoding ITF-2, a positive-acting transcription factor. Nucleic Acids Res 18:678.

Herbst A, Kolligs F. 2008. A conserved domain in the transcription factor ITF-2B attenuates its activity. Biochem Biophys Res Commun 370: 327–331.

Ince-Dunn G, Hall BJ, Hu SC, Ripley B, Huganir RL, Olson JM, Tapscott SJ, Ghosh A. 2006. Regulation of thalamocortical patterning and synaptic maturation by NeuroD2. Neuron 49:683–695.

Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, Martins RA, Kennedy BA, Hassell RG, Visser ME, et al. 2010. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet 42:684–687.

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, Geschwind DH, Mane SM, State MW, Sestan N. 2009. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron 62:494–509.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G, et al. 2011. Spatio-temporal transcriptome of the human brain. Nature 478:483–489.

Kolligs FT, Nieman MT, Winer I, Hu G, Van Mater D, Feng Y, Smith IM, Wu R, Zhai Y, Cho KR, et al. 2002. ITF-2, a downstream target of the Wnt/TCF pathway, is activated in human cancers with beta-catenin defects and promotes neoplastic transformation. Cancer Cell 1:145–155.

Kuot A, Hewitt AW, Griggs K, Klebe S, Mills R, Jhanji V, Craig JE, Sharma S, Burdon KP. 2012. Association of TCF4 and CLU polymorphisms with Fuchs' endothelial dystrophy and implication of CLU and TGFBI proteins in the disease process. Eur J Hum Genet 20:632–638.

Kurian SM, Le-Niculescu H, Patel SD, Bertram D, Davis J, Dike C, Yehyawi N, Lysaker P, Dustin J, Caligiuri M, et al. 2011. Identification of blood biomarkers for psychosis using convergent functional genomics. Mol Psychiatry 16:37–58.

Kwon E, Wang W, Tsai LH. 2011. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. Mol Psychiatry December 20. DOI: 10.1038/mp.2011.170. [Epub ahead of print]

la Cour T, Kiemer L, Molgaard A, Gupta R, Skriver K, Brunak S. 2004. Analysis and prediction of leucine-rich nuclear export signals. Protein Eng Des Sel 17:527–536.

Lang GK, Naumann GO. 1987. The frequency of corneal dystrophies requiring keratoplasty in Europe and the U.S.A. Cornea 6:209–211.

Laterza OF, Lim L, Garrett-Engele PW, Vlasakova K, Muniappa N, Tanaka WK, Johnson JM, Sina JF, Fare TL, Sistare FD, et al. 2009. Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clin Chem 55:1977–1983.

Lavon I, Zrihan D, Granit A, Einstein O, Fainstein N, Cohen MA, Zelikovitch B, Shoshan Y, Spektor S, Reubinoff BE, et al. 2010. Gliomas display a microRNA expression profile reminiscent of neural precursor cells. Neuro Oncol 12:422–433.

Lazorchak A, Jones ME, Zhuang Y. 2005. New insights into E-protein function in lymphocyte development. Trends Immunol 26:334–338.

Lewis DA, Gonzalez-Burgos G. 2008. Neuroplasticity of neocortical circuits in schizophrenia. Neuropsychopharmacology 33:141–165.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20.

Li Y, Li W, Yang Y, Lu Y, He C, Hu G, Liu H, Chen J, He J, Yu H. 2009. MicroRNA-21 targets LRRFIP1 and contributes to VM-26 resistance in glioblastoma multiforme. Brain Res 1286:13–18.

Li YJ, Minear MA, Rimmler J, Zhao B, Balajonda E, Hauser MA, Allingham RR, Eghrari AO, Riazuddin SA, Katsanis N, Gottsch JD, Gregory SG, Klintworth GK, Afshari NA. 2011. Replication of TCF4 through association and linkage studies in late-onset Fuchs endothelial corneal dystrophy. PLoS One 6(4):e18044.

Lin CH, Tapscott SJ, Olson JM. 2006. Congenital hypothyroidism (cretinism) in neuroD2-deficient mice. Mol Cell Biol 26:4311–4315.

Liu Y, Ray SK, Yang XQ, Luntz-Leybman V, Chiu IM. 1998. A splice variant of E2-2 basic helix-loop-helix protein represses the brain-specific fibroblast growth factor 1 promoter through the binding to an imperfect E-box. J Biol Chem 273:19269–19276.

Liu DZ, Tian Y, Ander BP, Xu H, Stamova BS, Zhan X, Turner RJ, Jickling G, Sharp FR. 2010. Brain and blood microRNA expression profiling of ischemic stroke, intracerebral hemorrhage, and kainate seizures. J Cereb Blood Flow Metab 30:92–101.

Lu R, Qu Y, Ge J, Zhang L, Su Z, Pflugfelder SC, Li DQ. 2012. Transcription factor TCF4 maintains the properties of human corneal epithelial stem cells. Stem Cells 30(4):753–761.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 104(Suppl1): 8597–8604.

Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, Barnes MR, Larminie C, Jones N, Lennon M, et al. 2009. Analysis of gene expression in

two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Mol Psychiatry 14:1083–1094.

McInnis MG, Swift-Scanlanl T, Mahoney AT, Vincent J, Verheyen G, Lan TH, Oruc L, Riess O, Van Broeckhoven C, Chen H, et al. 2000. Allelic distribution of CTG18.1 in Caucasian populations: Association studies in bipolar disorder, schizophrenia, and ataxia. Mol Psychiatry 5:439–442.

Meek KM, Leonard DW, Connon CJ, Dennis S, Khan S. 2003. Transparency, swelling and scarring in the corneal stroma. Eye (Lond) 17: 927–936.

Merz K, Herold S, Lie DC. 2011. CREB in adult neurogenesis-master and partner in the development of adult-born neurons? Eur J Neurosci 33:1078–1086.

Miyaoka T, Seno H, Ishino H. 1999. Increased expression of Wnt-1 in schizophrenic brains. Schizophr Res 38:1–6.

Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ, Luo S, Zhang L, van Velkinburgh JC, Farmer AD, et al. 2008. Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. PLoS ONE 3(11):e3625.

Muir T, Sadler-Riggleman I, Stevens JD, Skinner MK. 2006. Role of the basic helix-loop-helix protein ITF2 in the hormonal regulation of Sertoli cell differentiation. Mol Reprod Dev 73:491–500.

Murre C. 2005. Helix-loop-helix proteins and lymphocyte development. Nat Immunol 6(11):1079–1086.

Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, et al. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 58:537–544.

Nagasawa M, Schmidlin H, Hazekamp MG, Schotte R, Blom B. 2008. Development of human plasmacytoid dendritic cells depends on the combined action of the basic helix-loop-helix factor E2-2 the Ets factor Spi-B. Eur J Immunol 38:2389–2400.

Narayan S, Tang B, Head SR, Gilmartin TJ, Sutcliffe JG, Dean B, Thomas EA. 2008. Molecular profiles of schizophrenia in the CNS at different stages of illness. Brain Res 1239:235–248.

Nass D, Rosenwald S, Meiri E, Gilad S, Tabibian-Keissar H, Schlosberg A, Kuker H, Sion-Vardy N, Tobar A, Kharenko O, et al. 2009. MiR-92b and miR-9/9* are specifically expressed in brain primary tumors and can be used to differentiate primary from metastatic brain tumors. Brain Pathol 19:375–383.

Noda T, Kawamura R, Funabashi H, Mie M, Kobatake E. 2006. Transduction of NeuroD2 protein induced neural cell differentiation. J Biotechnol 126:230–236.

Northcott PA, Fernandez LA, Hagan JP, Ellison DW, Grajkowska W, Gillespie Y, Grundy R, Van Meter T, Rutka JT, Croce CM, et al. 2009. The miR-17/92 polycistron is up-regulated in sonic hedgehog-driven medulloblastomas and induced by N-myc in sonic hedgehog-treated cerebellar neural precursors. Cancer Res 69:3249–3255.

O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, International Schizophrenia Consortium, Corvin A. 2011. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Mol Psychiatry 16(3):286–292.

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. 2008. Functional organization of the transcriptome in human brain. Nat Neurosci 11:1271–1282.

Olson JM, Asakura A, Snider L, Hawkes R, Strand A, Stoeck J, Hallahan A, Pritchard J, Tapscott SJ. 2001. NeuroD2 is necessary for development and survival of central nervous system neurons. Dev Biol 234: 174–187.

Orrico A, Galli L, Zappella M, Lam CW, Bonifacio S, Torricelli F, Hayek G. 2001. Possible case of Pitt-Hopkins syndrome in sibs. Am J Med Genet 103(2):157–159.

Othman A, Frim DM, Polak P, Vujicic S, Arnason BG, Boullerne AI. 2011. Olig1 is expressed in human oligodendrocytes during maturation and regeneration. Glia 59:912–926.

Ozsolak F, Milos PM. 2011. RNA sequencing: Advances, challenges and opportunities. Nat Rev Genet 12:87–98.

Panman L, Andersson E, Alekseenko Z, Hedlund E, Kee N, Mong J, Uhde CW, Deng Q, Sandberg R, Stanton LW, Ericson J, Perlmann T. 2011. Transcription factor-induced lineage selection of stem-cell-derived neural progenitor cells. Cell Stem Cell 8:663–675.

Peippo MM, Simola KO, Valanne LK, Larsen AT, Kahkonen M, Auranen MP, Ignatius J. 2006. Pitt-Hopkins syndrome in two patients and further definition of the phenotype. Clin Dysmorphol 15:47–54.

Pemberton LF, Paschal BM. 2005. Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic 6:187–198.

Perkins DO, Jeffries CD, Jarskog LF, Thomson JM, Woods K, Newman MA, Parker JS, Jin J, Hammond SM. 2007. microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. Genome Biol 8:R27.

Persson P, Jogi A, Grynfeld A, Pahlman S, Axelson H. 2000. HASH-1 and E2-2 are expressed in human neuroblastoma cells and form a functional complex. Biochem Biophys Res Commun 274:22–31.

Petropoulos H. 2000. Analysis of the inhibition of MyoD activity by ITF-2B and full-length E12/E47. J Biol Chem 275:25095–25101.

Pickard BS, Millar JK, Porteous DJ, Muir WJ, Blackwood DH. 2005. Cytogenetics and gene discovery in psychiatric disorders. Pharmacogenomics J 5:81–88.

Pidsley R, Mill J. 2011. Epigenetic studies of psychosis: Current findings, methodological approaches, and implications for postmortem research. Biol Psychiatry 69:146–156.

Pierson J, Hostager B, Fan R, Vibhakar R. 2008. Regulation of cyclin dependent kinase 6 by microRNA 124 in medulloblastoma. J Neurooncol 90:1–7.

Pitt D, Hopkins I. 1978. A syndrome of mental retardation, wide mouth and intermittent overbreathing. Aust Paediatr J 14:182–184.

Poethig RS. 2009. Small RNAs and developmental timing in plants. Curr Opin Genet Dev 19:374–378.

Pscherer A, Dorflinger U, Kirfel J, Gawlas K, Ruschoff J, Buettner R, Schule R. 1996. The helix-loop-helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin receptor II gene. EMBO J 15:6680–6690.

Quednow BB, Ettinger U, Mössner R, Rujescu D, Giegling I, Collier DA, Schmechtig A, Kühn KU, Möller HJ, Maier W, Wagner M, Kumari V. 2011. The schizophrenia risk allele C of the TCF4 rs9960767 polymorphism disrupts sensorimotor gating in schizophrenia spectrum and healthy volunteers. J Neurosci 31:6684–6691.

Quednow BB, Brinkmeyer J, Mobascher A, Nothnagel M, Musso F, Gründer G, Savary N, Petrovsky N, Frommann I, Lennertz L, Spreckelmeyer KN, Wienker TF, Dahmen N, Thuerauf N, Clepce M, Kiefer F, Majic T, Mössner R, Maier W, Gallinat J, Diaz-Lacava A, Toliat MR, Thiele H, Nürnberg P, Wagner M, Winterer G. 2012. Schizophrenia risk polymorphisms in the TCF4 gene interact with smoking in the modulation of auditory sensory gating. Proc Natl Acad Sci USA 109: 6271–6276.

Quong MW, Romanow WJ, Murre C. 2002. E protein function in lymphocyte development. Annu Rev Immunol 20:301–322.

Ravanpay AC, Olson JM. 2008. E protein dosage influences brain development more than family member identity. J Neurosci Res 86:1472–1481.

Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, et al. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140: 744–752.

Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, Lin DY, Duan J, Ophoff RA, Andreassen OA, et al. 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43: 969–976.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4:651–657.

Ross SE, Greenberg ME, Stiles CD. 2003. Basic helix-loop-helix factors in cortical development. Neuron 39:13–25.

Rossbach M. 2011. Non-coding RNAs in neural networks, REST-assured. Front Genet 2:8.

Rozowsky J, Euskirchen G, Auerbach RK, Zhang ZD, Gibson T, Bjornson R, Carriero N, Snyder M, Gerstein MB. 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol 27:66–75.

Rujescu D, Collier DA. 2009. Dissecting the many genetic faces of schizophrenia. Epidemiol Psichiatr Soc 18:91–95.

Saarikettu J. 2004. Calcium/calmodulin inhibition of transcriptional activity of E-proteins by prevention of their binding to DNA. J Biol Chem 279:41004–41011.

Saba R, Goodman CD, Huzarewich RL, Robertson C, Booth SA. 2008. A miRNA signature of prion induced neurodegeneration. PLoS ONE 3(11):e3652.

Sepp M, Kannike K, Eesmaa A, Urb M, Timmusk T. 2011. Functional diversity of human basic helix-loop-helix transcription factor TCF4 isoforms generated by alternative 5' exon usage and splicing. PLoS ONE 6:e22138.

Sethupathy P, Borel C, Gagnebin M, Grant GR, Deutsch S, Elton TS, Hatzigeorgiou AG, Antonarakis SE. 2007. Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3′ untranslated region: A mechanism for functional single-nucleotide polymorphisms related to phenotypes. Am J Hum Genet 81:405–413.

Silber J, Lim DA, Petritsch C, Persson AI, Maunakea AK, Yu M, Vandenberg SR, Ginzinger DG, James CD, Costello JF, et al. 2008. miR-124 and miR-137 inhibit proliferation of glioblastoma multiforme cells and induce differentiation of brain tumor stem cells. BMC Med 6:14.

Singh HA. 1993. Mental retardation, macrostomia and hyperpnoea syndrome. J Paediatr Child Health 29:156–157.

Skerjanc IS, Truong J, Filion P, McBurney MW. 1996. A splice variant of the ITF-2 transcript encodes a transcription factor that inhibits MyoD activity. J Biol Chem 271:3555–3561.

Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, Edenberg HJ, Nurnberger JI Jr, Rietschel M, Blackwood D, et al. 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43:977–983.

Sobrado VR, Moreno-Bueno G, Cubillo E, Holt LJ, Nieto MA, Portillo F, Cano A. 2009. The class I bHLH factors E2-2A and E2-2B regulate EMT. J Cell Sci 122:1014–1024.

Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietilainen OP, Mors O, Mortensen PB, et al. 2009. Common variants conferring risk of schizophrenia. Nature 460:744–747.

Steinberg S, de Jong S, Andreassen OA, Werge T, Borglum AD, Mors O, Mortensen PB, Gustafsson O, Costas J, Pietilainen OP, et al. 2011. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. Hum Mol Genet 20:4076–4081.

Stephan KE, Friston KJ, Frith CD. 2009. Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bull 35:509–527.

Talebizadeh Z, Butler MG, Theodoro MF. 2008. Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. Autism Res 1:240–250.

Thierry-Mieg D, Thierry-Mieg J. 2006. AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biol 7(Suppl1): S12–S14.

Uziel T, Karginov FV, Xie S, Parker JS, Wang YD, Gajjar A, He L, Ellison D, Gilbertson RJ, Hannon G, et al. 2009. The miR-17∼92 cluster collaborates with the Sonic Hedgehog pathway in medulloblastoma. Proc Natl Acad Sci USA 106:2812–2817.

Van Balkom ID, Quartel S, Hennekam RC. 1998. Mental retardation, "coarse" face, and hyperbreathing: Confirmation of the Pitt-Hopkins syndrome. Am J Med Genet 75:273–276.

van Os J, Kapur S. 2009. Schizophrenia. Lancet 374:635–645.

Vawter MP. 2000. Dysregulation of the neural cell adhesion molecule and neuropsychiatric disorders. Eur J Pharmacol 405:385–395.

Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. 2008. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 4:e1000214.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA enhancer browser—A database of tissue-specific human enhancers. Nucleic Acids Res 35:D88–D92.

Wang VY, Rose MF, Zoghbi HY. 2005. Math1 expression redefines the rhombic lip derivatives and reveals novel lineages within the brainstem and cerebellum. Neuron 48:31–43.

Watowich SS, Liu YJ. 2010. Mechanisms regulating dendritic cell specification and development. Immunol Rev 238:76–92.

Whalen S, Heron D, Gaillon T, Moldovan O, Rossi M, Devillard F, Giuliano F, Soares G, Mathieu-Dramard M, Afenjar A, et al. 2012. Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: Clinical score and further delineation of the TCF4 mutational spectrum. Hum Mutat 33:64–72.

Wikstrom I, Forssell J, Penha-Goncalves MN, Bergqvist I, Holmberg D. 2008. A role for E2-2 at the DN3 stage of early thymopoiesis. Mol Immunol 45:3302–3311.

Williams HJ, Craddock N, Russo G, Hamshere ML, Moskvina V, Dwyer S, Smith RL, Green E, Grozeva D, Holmans P, et al. 2011a. Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. Hum Mol Genet 20:387–391.

Williams HJ, Moskvina V, Smith RL, Dwyer S, Russo G, Owen MJ, O'Donovan MC. 2011b. Association between TCF4 and schizophrenia does not exert its effect by common nonsynonymous variation or by influencing cis-acting regulation of mRNA expression in adult human brain. Am J Med Genet Part B Neuropsychiatr Genet 156B:781–784.

Yoon SO, Chikaraishi DM. 1994. Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. J Biol Chem 269:18453–18462.

Zaghloul NA, Katsanis N. 2010. Functional modules, mutational load and human genetic disease. Trends Genet 26:168–176.

Zhang C. 2009. MicroRNA and vascular smooth muscle cell phenotype: New therapy for atherosclerosis? Genome Med 1(9):85.

Zierhut K, Bogerts B, Schott B, Fenker D, Walter M, Albrecht D, Steiner J, Schutze H, Northoff G, Duzel E, et al. 2010. The role of hippocampus dysfunction in deficient memory encoding and positive symptoms in schizophrenia. Psychiatry Res 183:187–194.

Zweier C. 2007. Haploinsufficiency of TCF4 causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). Am J Hum Genet 80:994–1001.

Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MAW, Ekici AB, Reis A. 2009. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in drosophila. Am J Hum Genet 85:655–666.

SUPPLEMENTARY TABLES

Supplementary Table I. Output from the TargetPserver, tested on TCF4 sequences (four sequences from Human and one from Chimpanzee). The "Loc" column contains the prediction of N-terminal sorting signal: "S" for secretory.

```
### targetp v1.1 prediction results ###################################
Number of query sequences:  5
Cleavage site predictions included.
Using NON-PLANT networks.

Name                   Len            mTP     SP  other  Loc  RC  TPlen
--------------------------------------------------------------------------
Chimpanzee             752          0.041  0.856  0.122   S    2     27
ENSP00000381382        773          0.041  0.858  0.114   S    2     31
ENSP00000348374        667          0.106  0.038  0.896   _    2      -
ENSP00000346440        671          0.106  0.038  0.896   _    2      -
ENSP00000409447        511          0.091  0.094  0.893   _    2      -
--------------------------------------------------------------------------
cutoff                              0.000  0.000  0.000
```

Supplementary Table II Multi-species conserved sequences (MCS) in the comparison of human and mouse TCF4 genes.

start	end	Length (bp)	Identity (%)	Region	SNP
51049707	51049893	190	71.60	intron 17	
51050347	51050466	125	72.00	intron 17	
51050514	51050737	224	80.80	intron 17	
51050991	51051461	474	85.00	intron 16	
51052323	51052683	377	76.90	intron 16	
51053084	51053695	665	84.50	intron 15	
51055983	51056095	113	71.70	intron 15	rs10515969, rs7237259
51057819	51058316	501	83.60	intron 15	
51058549	51058776	229	72.50	intron 15	
51061146	51061260	118	70.30	intron 15	
51061438	51061818	392	82.40	intron 15	
51062110	51062428	335	71.60	intron 15	
51062616	51063330	717	76.70	intron 15	
51063434	51063696	265	85.30	intron 15	
51064989	51065099	114	70.20	intron 15	
51065437	51065572	138	68.80	intron 15	
51067615	51067721	107	69.20	intron 15	
51068360	51068476	117	68.40	intron 15	
51069305	51069423	119	70.60	intron 15	
51069578	51070192	626	71.90	intron 15	
51070216	51070332	124	75.00	intron 15	
51070797	51071449	674	81.00	intron 15	
51072492	51072588	98	71.40	intron 15	
51073030	51073351	327	75.50	intron 14	
51073030	51073351	327	75.50	intron 14	
51073923	51074017	96	70.80	intron 14	
51073923	51074017	96	70.80	intron 14	
51074094	51074290	202	77.20	intron 14	
51074960	51075093	134	72.40	intron 14	
51075309	51075410	102	70.60	intron 14	
51076233	51076561	339	80.50	intron 13	
51077556	51077727	172	71.50	intron 13	
51079589	51079694	108	87.00	intron 12	
51079589	51079694	108	87.00	intron 12	
51079763	51080040	285	86.70	intron 11	
51082378	51082818	446	76.70	intron 11	
51083072	51083458	390	71.80	intron 11	
51083508	51083619	116	71.60	intron 11	
51083686	51083817	135	71.10	intron 11	
51083881	51084313	471	72.20	intron 11	
51085262	51085700	453	76.40	intron 11	
51086019	51086116	98	71.40	intron 11	
51086372	51086853	490	78.40	intron 11	
51087603	51087706	104	71.20	intron 11	

51087779	51087903	129	69.80	intron 11	
51088271	51088381	113	71.70	intron 10	
51090237	51090575	356	71.60	intron 10	
51091211	51091426	218	74.30	intron 10	
51091655	51091956	307	75.90	intron 10	
51092437	51092597	161	71.40	intron 10	
51093003	51093139	138	71.00	intron 10	
51093516	51093772	267	70.80	intron 10	
51093982	51094199	226	79.60	intron 9	
51095494	51095980	503	76.30	intron 9	
51096138	51096373	241	69.70	intron 9	
51098382	51098997	637	78.30	intron 8	
51099001	51099258	264	75.80	intron 8	
51104202	51104301	101	70.30	intron 8	
51107153	51107366	214	72.90	intron 8	
51107382	51107644	266	75.90	intron 8	
51107700	51108040	348	71.80	intron 8	
51108205	51108313	112	69.60	intron 8	
51109523	51109936	436	77.10	intron 8	
51110274	51110405	132	68.20	intron 8	
51110474	51110647	181	69.10	intron 8	
51111275	51111373	100	70.00	intron 8	
51111492	51111640	149	67.10	intron 8	
51114275	51114582	312	80.80	intron 8	
51114809	51115438	638	80.30	intron 8	
51115461	51115560	105	71.40	intron 8	
51116514	51116706	196	69.90	intron 8	
51116867	51116969	104	75.00	intron 8	
51118287	51118437	153	69.30	intron 8	
51118478	51118603	129	79.80	intron 8	
51118624	51119442	862	80.70	intron 8	
51119469	51120022	576	79.00	intron 8	
51120105	51120371	270	78.10	intron 8	
51120572	51121022	454	78.60	intron 8	
51123111	51123455	353	90.10	intron 8	
51123456	51123584	131	78.60	intron 8	
51123685	51123888	205	74.10	intron 8	
51124086	51124191	107	74.80	intron 8	
51124422	51125319	916	87.40	intron 8	
51125402	51126023	632	74.80	intron 8	
51128561	51128810	267	69.70	intron 8	
51129197	51129329	137	73.70	intron 8	
51129690	51129819	132	69.70	intron 8	
51130012	51130361	351	74.10	intron 8	
51130717	51130944	233	73.40	intron 8	
51130984	51131688	725	75.40	intron 8	
51132275	51132825	587	77.90	intron 8	
51132831	51132946	128	69.50	intron 8	

51132949	51133359	428	73.80	intron 8	
51134184	51134279	104	70.20	intron 8	
51134417	51134515	99	74.70	intron 8	
51135185	51135415	243	74.90	intron 8	
51135573	51135723	151	69.50	intron 8	
51136091	51136459	373	81.00	intron 8	
51136718	51136867	155	71.60	intron 8	
51136993	51137369	388	75.50	intron 8	
51137403	51138191	790	80.00	intron 8	
51138887	51139597	738	73.20	intron 8	
51139750	51140305	561	88.10	intron 8	
51140358	51140843	504	85.10	intron 8	
51141005	51141113	114	71.90	intron 8	
51141180	51141336	157	73.20	intron 8	
51141350	51142013	666	82.00	intron 8	
51142085	51142203	120	71.70	intron 8	
51142358	51142691	340	74.40	intron 8	
51144220	51144354	136	69.90	intron 8	
51144980	51145086	112	73.20	intron 8	
51145125	51145204	85	75.30	intron 8	
51145747	51145910	165	77.00	intron 8	
51146634	51146771	138	81.90	intron 8	
51147026	51147275	258	72.90	intron 8	
51150277	51150374	100	70.00	intron 8	
51150663	51150821	159	67.30	intron 8	
51151405	51151541	139	71.20	intron 8	
51151552	51151647	100	72.00	intron 8	
51151909	51152141	235	68.90	intron 8	
51152320	51152429	114	71.10	intron 8	
51152811	51153007	198	73.20	intron 8	
51153050	51153754	712	80.30	intron 8	
51154025	51154204	180	75.00	intron 8	
51154836	51155456	625	94.90	intron 8	
51155458	51155940	489	77.70	intron 8	
51156127	51156242	134	76.90	intron 8	
51156245	51156722	480	84.40	intron 8	
51156745	51156862	118	69.50	intron 8	
51157156	51157491	339	95.90	intron 8	
51157560	51157675	119	77.30	intron 8	
51157934	51158049	116	72.40	intron 8	
51158348	51158454	107	70.10	intron 8	
51160581	51160677	102	70.60	intron 8	
51160904	51161329	431	79.10	intron 8	
51161420	51161523	104	71.20	intron 8	
51162698	51162802	113	73.50	intron 8	
51164486	51164582	101	70.30	intron 8	
51165167	51165261	100	70.00	intron 8	
51167096	51167261	166	69.90	intron 8	

51167280	51167733	470	79.10	intron 8	
51167805	51167892	88	77.30	intron 8	
51168126	51168435	312	84.90	intron 7	
51168488	51168587	103	70.90	intron 7	
51169233	51169359	131	70.20	intron 6	
51169876	51170040	177	74.60	intron 6	
51170043	51170582	540	95.90	intron 6	
51171081	51171198	123	73.20	intron 6	
51171271	51171388	118	68.60	intron 6	
51171440	51171583	148	69.60	intron 6	
51171673	51171952	285	78.90	intron 6	
51172050	51172178	130	71.50	intron 6	
51172676	51173038	365	82.20	intron 6	
51173075	51173949	890	79.00	intron 6	
51174258	51174357	101	70.30	intron 6	
51175041	51175361	322	74.20	intron 6	
51175501	51175601	101	70.30	intron 6	
51175789	51175971	189	72.00	intron 6	
51177714	51177813	100	70.00	intron 6	
51180080	51180197	118	68.60	intron 6	
51180838	51181188	353	73.10	intron 6	
51185607	51185726	120	70.80	intron 6	rs8092679
51187562	51187718	158	69.00	intron 6	
51187839	51188132	310	76.10	intron 6	
51196322	51196461	149	65.80	intron 6	
51200933	51201074	142	68.30	intron 6	rs9646596, rs17594301, rs1788025
51202334	51202632	303	73.30	intron 6	
51204105	51204224	122	69.70	intron 6	rs9950000
51204271	51204368	99	70.70	intron 6	
51205030	51205162	135	69.60	intron 6	rs9958125, rs9320010
51205471	51206107	641	81.90	intron 6	
51208087	51208186	100	71.00	intron 6	
51209211	51209429	220	73.20	intron 6	rs17594526
51212022	51212529	512	82.20	intron 6	
51212754	51212878	125	76.80	intron 6	
51213962	51214190	233	75.50	intron 6	
51217758	51217894	137	73.00	intron 6	rs11152369, rs1759472, rs17594665, rs17089778
51218083	51218146	64	98.40	intron 6	
51218413	51218633	225	72.90	intron 6	rs17509991
51219030	51219467	446	74.00	intron 6	
51219582	51219674	94	71.30	intron 6	
51219749	51220491	776	79.60	intron 6	rs17510124
51221076	51221564	501	76.80	intron 6	rs2958178
51221748	51222283	556	82.70	intron 5	
51222704	51222816	115	70.40	intron 5	
51224012	51224634	631	86.10	intron 5	
51224693	51225473	796	76.10	intron 5	
51225584	51225746	163	76.70	intron 5	

51225765	51225842	97	71.10	intron 5	
51226459	51226663	207	74.40	intron 5	
51226883	51226989	107	72.00	intron 5	rs10503001
51227021	51227494	481	78.40	intron 5	rs17089789
51229257	51229614	359	85.00	intron 5	
51233611	51233787	185	68.10	intron 5	
51233818	51233910	93	71.00	intron 5	
51234579	51234670	93	72.00	intron 5	
51234841	51235407	580	82.20	intron 5	
51236238	51236559	324	74.70	intron 5	rs11873788, rs17089803
51237168	51237375	209	72.70	intron 5	
51237773	51238363	617	69.90	intron 5	
51238966	51239112	159	74.20	intron 5	
51239154	51239273	120	83.30	intron 5	
51239661	51239837	180	70.00	intron 5	
51239911	51240028	121	70.20	intron 5	
51240164	51240328	168	71.40	intron 5	
51240348	51241258	954	83.80	intron 5	
51241259	51241502	248	86.30	intron 5	
51241503	51242386	908	81.10	intron 5	
51242493	51242628	136	72.80	intron 5	
51243191	51243288	99	70.70	intron 5	
51243388	51243586	203	78.80	intron 5	
51243620	51243717	105	75.20	intron 5	
51243775	51244291	534	79.80	intron 5	
51244371	51244954	593	80.30	intron 5	
51245611	51245718	108	71.30	intron 5	
51248095	51248211	117	70.10	intron 5	rs8086206, rs9320016
51248297	51248392	100	70.00	intron 5	
51254304	51254636	340	73.80	intron 5	
51255339	51255477	140	68.60	intron 5	
51255651	51255811	172	68.00	intron 5	
51258583	51258684	102	70.60	intron 5	rs17595731
51258721	51258819	99	70.70	intron 5	
51259370	51259799	432	77.30	intron 5	
51260184	51260941	772	77.20	intron 5	
51261362	51261489	129	69.80	intron 5	
51262263	51262412	150	68.70	intron 5	
51262940	51263057	121	70.20	intron 5	
51265920	51266217	308	73.40	intron 5	
51266507	51266822	321	74.50	intron 5	
51266952	51267159	212	75.00	intron 5	
51267676	51267892	219	70.30	intron 5	
51267983	51268148	166	71.70	intron 5	
51270389	51270573	187	73.80	intron 5	
51271459	51271575	118	72.00	intron 5	
51274038	51274192	157	71.30	intron 5	
51274209	51274566	363	84.60	intron 5	

51274769	51274920	153	70.60	intron 5	
51276229	51276546	318	80.80	intron 5	
51276859	51277448	597	78.60	intron 5	
51278230	51278668	449	71.70	intron 5	rs17089826
51280222	51280342	123	74.80	intron 4	
51280913	51281154	246	73.20	intron 4	
51282367	51282590	224	92.90	intron 3	
51282591	51282763	178	74.20	intron 3	
51283334	51283636	304	76.60	intron 3	
51283678	51283923	246	69.90	intron 3	
51284903	51285223	330	78.80	intron 3	
51285287	51285386	104	73.10	intron 3	
51285451	51285566	117	71.80	intron 3	
51287955	51288241	301	71.80	intron 3	rs17511376
51288842	51288962	125	69.60	intron 3	
51289207	51289513	308	85.70	intron 3	
51291454	51291598	150	75.30	intron 3	
51291798	51292283	486	86.00	intron 3	rs17596267
51295394	51295499	106	74.50	intron 3	rs12326678
51296229	51296854	639	79.20	intron 3	
51297956	51298330	377	75.10	intron 3	rs17511755, rs8090422, rs8086784
51302043	51302389	362	71.30	intron 3	
51302393	51302504	114	71.10	intron 3	
51304125	51304288	167	70.70	intron 3	
51305937	51306112	178	71.30	intron 3	rs9960767
51306134	51306261	128	88.30	intron 3	
51306263	51306481	239	74.50	intron 3	
51306482	51306616	139	71.20	intron 3	
51307015	51307114	100	70.00	intron 3	
51308104	51308216	113	68.10	intron 3	rs2924333
51308225	51308332	109	72.50	intron 3	
51309043	51309143	101	71.30	intron 3	
51309206	51310328	1136	78.50	intron 3	
51312137	51312230	96	72.90	intron 3	rs2958187
51312573	51312885	314	79.90	intron 3	
51313088	51313346	261	75.10	intron 3	
51313593	51313855	270	74.10	intron 3	
51314366	51314663	299	80.90	intron 3	
51314673	51315159	495	70.50	intron 3	
51315256	51315406	152	71.70	intron 3	
51315418	51315800	387	76.50	intron 3	
51315830	51315952	124	71.80	intron 3	
51316029	51316395	373	74.80	intron 3	
51317454	51317608	158	74.10	intron 3	rs17596974
51318524	51319116	609	79.60	intron 3	rs12327270, rs2958189
51319413	51320056	661	73.10	intron 3	
51320753	51320877	125	68.80	intron 3	
51320957	51321301	353	74.20	intron 3	

51321800	51321895	96	70.80	intron 3	
51322999	51323647	657	83.10	intron 3	rs2958158
51324689	51324784	96	70.80	intron 3	
51326723	51326857	138	70.30	intron 3	rs2924338
51326928	51327053	127	71.70	intron 3	
51327069	51327199	135	72.60	intron 3	
51327283	51327423	142	69.00	intron 3	
51328072	51328251	182	69.20	intron 3	
51328713	51328905	195	71.30	intron 3	
51329036	51329354	322	85.70	intron 3	
51329356	51329551	196	71.90	intron 3	
51331730	51331936	209	80.90	intron 3	
51332038	51332540	506	72.90	intron 3	
51336093	51336242	150	68.00	intron 3	rs12606243, rs624244
51338285	51338917	647	81.00	intron 3	
51339171	51339244	74	82.40	intron 3	
51339296	51339745	471	70.90	intron 3	
51342793	51343187	398	82.20	intron 3	
51345824	51346201	397	71.30	intron 3	rs10401120
51346513	51346622	110	71.80	intron 3	
51347795	51347953	161	70.80	intron 3	
51348404	51348976	579	82.60	intron 3	
51350709	51350817	109	69.70	intron 3	
51351270	51351379	115	73.00	intron 3	
51351484	51351769	288	82.30	intron 3	
51351824	51351919	96	71.90	intron 3	
51352019	51352113	100	72.00	intron 3	
51353502	51353601	100	70.00	intron 3	
51353752	51353848	101	70.30	intron 3	
51353901	51354358	472	71.40	intron 3	
51354600	51354837	241	69.70	intron 3	
51356319	51356538	223	80.70	intron 3	
51356636	51356735	100	71.00	intron 3	
51357587	51357817	235	74.00	intron 3	rs17597926
51359367	51359610	249	74.30	intron 3	
51359698	51359797	100	71.00	intron 3	
51359912	51360009	100	70.00	intron 3	
51360249	51360339	94	72.30	intron 3	
51361436	51362227	801	83.40	intron 3	
51362427	51362531	105	72.40	intron 3	
51363351	51363456	109	69.70	intron 3	
51364846	51365123	282	86.20	intron 3	
51365412	51365556	148	69.60	intron 3	
51365958	51366251	307	69.40	intron 3	
51366319	51366708	401	79.80	intron 3	
51368799	51369023	229	80.30	intron 3	
51369581	51369684	105	71.40	intron 3	
51370876	51371166	295	81.70	intron 3	

51373371	51373618	250	71.60	intron 3	
51373812	51373949	140	73.60	intron 3	
51373954	51374456	504	91.70	intron 3	
51377314	51377813	502	79.70	intron 3	
51377977	51378356	395	75.70	intron 3	
51380267	51380399	134	72.40	intron 3	
51380443	51380546	106	71.70	intron 3	
51392245	51392430	186	70.40	intron 3	
51394439	51394534	99	70.70	intron 3	
51395534	51395816	296	71.60	intron 3	
51395843	51396467	635	80.30	intron 3	
51396822	51397329	516	76.90	intron 3	
51397474	51397687	230	70.90	intron 3	
51397734	51398621	907	80.30	intron 3	
51398696	51398969	281	71.50	intron 3	
51400459	51400937	479	79.30	intron 3	
51401085	51401219	135	68.90	intron 3	
51401326	51401499	176	71.00	intron 3	
51401752	51401863	119	69.70	intron 3	
51402302	51402699	400	71.50	intron 3	
51402713	51403034	334	71.90	intron 3	
51403062	51403255	196	69.90	intron 3	
51403342	51403508	174	74.70	intron 3	
51403582	51403731	150	79.30	intron 2	
51403815	51404078	272	76.50	intron 2	
51404129	51404341	219	80.40	intron 2	
51404452	51404761	322	75.20	intron 2	
51405004	51405273	303	72.90	intron 2	
51405366	51405366	502	80.70	intron 1	
51405840	51405974	139	68.30	intron 1	
		9866600.00	74.82		

Supplementary Table III. MCS conserved between human and mouse from vista enhancer project

ID	Human (hg19)		Mouse (mm9)		Expression
	Coordinates	Bracketing Genes	Coordinates	Bracketing Genes	
hs376	chr18:53,089,625-53,091,079	TCF4(intragenic)	chr18:69,658,815-69,660,451	Tcf4(intragenic)	
hs713	chr18:53,530,668-53,531,881	TCF4-FLJ45743	chr18:69,284,214-69,285,465	Mc2r-Tcf4	
hs744	chr18:53,567,879-53,568,822	TCF4-FLJ45743	chr18:69,255,489-69,256,442	Mc2r-Tcf4	
hs1561	chr18:52,971,887-52,975,873	TCF4(intragenic)	chr18:69,766,085-69,769,703	Tcf4(intragenic)	

Supplementary Table IV Genetic variations within the Human element [hs376].

chromStart	chromEnd	name	class	valid	func	locType
51240647	51240648	rs2958170	single	by-cluster,by-2hit-2allele,by-1000genomes	intron	exact
51241299	51241299	rs35586642	insertion	unknown	intron	between
51241402	51241402	rs10623644	mixed	unknown	intron	between
51241403	51241403	rs33992145	mixed	unknown	intron	between
51241411	51241411	rs56100102	insertion	unknown	intron	between
51241412	51241412	rs67607050	insertion	unknown	intron	between
51241646	51241648	rs71875260	deletion	unknown	intron	range
51241654	51241658	rs35538327	deletion	unknown	intron	range
51241655	51241656	rs2919445	single	by-cluster,by-2hit-2allele	intron	exact
51241655	51241659	rs34425939	deletion	unknown	intron	range
51241655	51241657	rs56336951	deletion	unknown	intron	range
51241657	51241658	rs72926988	single	by-1000genomes	intron	exact
51242057	51242058	rs9959527	single	by-hapmap	intron	exact

Supplementary Table V. Expression pattern of TCF4 in human brain. Data were retrieved from Sestan Lab Human Brain Atlas Microarrays data set.

Name	neocortex oPFC	neocortex mPFC	neocortex dlPFC	neocortex vlPFC	neocortex motorcensory	neocortex parietal	neocortex tempassociation	neocortex tempauditory	neocortex occip	hippocamp	striatum	thalamus	cerebellum
3808866	-0.1147891	0.3865555	0.3972245	0.20546	0.277334	0.6795895	0.4054485	0.683712	0.5902855	0.570667	-1.689625	-3.33738	-1.09667
3789066	-0.7840035	0.379119	-0.010164365	-0.116527132	0.01611175	1.0867975	-0.0662543	0.223243	-0.10121085	0.4434425	-0.498906	-0.6455635	-0.452033
3808867	-0.0112425	0.2278435	0.1331725	0.2639935	0.079251575	0.6706635	0.316758	0.471803	0.554426	0.703846	-2.12952	-3.50853	-1.10312
3789067	-0.249277	-0.09167295	0.13375945	0.06991595	0.17542895	0.3996075	-0.0318951	0.2446305	0.1225735	0.075307	-0.08143255	-0.14767115	-0.185426
3808868	-0.137106	0.224748	0.154433	0.17092525	0.1738825	0.5421775	0.332589	0.506473	0.399994	0.519147	-1.799105	-3.35227	-0.720093
3789069	-1.093585	-0.56134	0.02605628	-0.1838945	0.271227	1.035878	0.2963396	0.7112865	0.460352	0.1199413	-1.102575	-2.39781	-1.13424
3789060	-0.2922195	-0.13674255	-0.7909065	-0.2396695	-0.066666	0.5591155	-0.01534435	0.4145845	-0.265674	0.566293	-1.346855	-2.391105	-1.26876
3808869	-0.297582	0.10173005	0.1347474	0.184156975	0.0999615	0.444111	0.3869505	0.493711	0.516759	0.3569505	-1.440715	-2.980365	-0.633969
3789061	-0.42609273	1.19257	0.9542575	0.966764	1.64284	2.332915	0.883394	1.64854	1.858875	1.839495	-0.3759455	-1.24837	0.0957754
3808880	-0.02852195	0.1958895	0.07595221	0.0626304	0.187428	0.3777485	0.303292	0.2299005	0.3016985	0.3371755	-1.251365	-2.42444	-0.366282
3789063	0.014838	0.0788647	0.00962035	0.01015725	0.0468885	0.0176808	0.02682475	0.0494406	-0.012283585	0.00326591	-0.052824	-0.0604443	-0.058124
3808881	-0.132193885	0.06249955	0.06508594	0.03229593	0.0540526	0.2795265	0.01799255	0.204947	0.5297145	0.339189	-1.69877	-3.062575	-0.731118
3789064	-0.064383	0.4159005	0.1945854	0.2979515	0.785254	1.09176	0.4432375	0.7953415	0.6277535	0.822065	-0.385301	-1.57075	-0.56311
3808882	-0.132065	-0.0784081	-0.105761	-0.1949405	-0.07703055	0.29738	0.065598505	0.1166905	0.1587135	0.0881165	-1.78327	-3.166355	-0.692401
3808886	-0.0968593	-0.02056785	-0.10951125	-0.05092865	-0.02347535	0.202199	0.0703584	0.13521	0.2387395	0.2145145	-1.60545	-2.88346	-0.475001
3808868	-0.2166105	0.13455545	-0.0614913	-0.0245459	0.265118	0.2986165	0.215325	0.2719905	0.3587705	0.41053	-1.579805	-2.714	-0.568239
3808870	-0.1832225	-0.1302275	-0.1450193	-0.08877665	0.038989995	0.0563332	0.0564943	0.1331808	0.244405	0.1305847	-1.69412	-2.819325	-0.494351
3808871	-0.0699455	0.01118395	-0.054011	0.03216845	0.1210095	0.2088415	0.0800813	0.189639	0.262235	0.1933385	-1.26538	-2.44006	-0.50727
3808890	-0.0681075	-0.12177	-0.280958	-0.2123145	-0.112084	0.1351135	-0.05648435	0.0837006	0.128514	0.22791005	-1.94533	-3.027815	-0.66901
3808891	-0.336215	-0.252026	-0.3956535	-0.173504	-0.2067332	0.133509	0.0082668	-0.06555305	0.194083	0.16259175	-1.931715	-2.745565	-0.650283
3808887	-0.07673665	-0.1626531	-0.415509	-0.263953	-0.219911	0.13543695	-0.207877	-0.1734555	-0.03575455	-0.05769595	-1.77944	-3.071865	-0.664397
3789078	0.04319662	0.0126963	0.02932945	0.00439605	0.0322324	0.1492155	0.15038555	0.167554	0.0481379	0.010184985	0.014516925	0.004761265	0.145751
3789082	-0.14704435	-0.400501	0.604586	-0.216742	-0.1071172	0.1824709	0.166909	0.038899965	-0.5581695	-0.2997364	-0.7999145	-0.5546975	-0.177846
3808904	-0.1967805	-0.128835	-0.3197115	-0.178422	-0.09110792	0.0867386	0.06740115	-0.012082565	0.11081285	0.2340175	-1.498265	-2.549635	-0.57798
3808905	0.03475535	-0.0354956	-0.0381637	-0.1049325	0.04610925	0.0835927	0.10412665	0.1947845	0.2633245	0.325401	-1.46366	-2.75296	-0.504428
3808909	-0.13567155	-0.1918055	-0.1829125	-0.1715115	-0.05260605	0.01643703	0.02055848	0.10329115	0.1972985	0.163505	-1.667565	-2.77569	-0.575599
3808912	-0.0977691	0.00532125	0.0679903	-0.388405	-0.422168	-0.1445171	0.210845	0.0283744	0.2123305	0.510936	-1.518705	-3.249245	-0.133265
3808823	-0.4319645	-0.1553257	-0.405243	-0.284626	-0.1469395	0.0205702	0.174613	-0.09852545	0.0301306	0.176692	-2.058745	-2.75538	-0.918885
3808838	-0.149481	-0.432261	-0.4652775	-0.3894655	-0.3385675	-0.00644042	0.0997196	-0.10168735	0.0753065	0.1908749	-2.03221	-3.240855	-0.955575
3808840	-1.51568	0.239896	0.505082	-0.0403775	0.050976	1.11935	0.2327685	0.942648	0.6371185	1.728625	-1.800095	-2.333525	-0.931123
3808849	0.148069015	-0.163469	-0.2424865	-0.1809695	-0.1904295	0.2335155	0.3238305	0.287963	0.3355375	0.324298	-2.098785	-3.029795	-0.867044
3808860	-0.447522	-0.526392	-0.5468425	-0.565485	-0.5544775	-0.2318855	0.08999035	-0.0702541	-0.08973205	-0.1227765	-1.96344	-2.72464	-0.846468
3808872	-0.4015585	-0.2468795	0.2970475	-0.09367025	-0.3579956	0.654943	1.27712	1.048853	0.4273195	0.1718095	-1.533095	-2.295515	-0.371985
3808890	0.3144425	-0.182769	-0.712688	0.1726782	-0.083741	0.6859455	0.1370545	0.170923	0.2003379	0.485921	-2.274185	-3.138635	-1.05028
3789178	0.50543	-0.5769695	-0.294673	-0.21474644	-0.1924782	-0.143441405	0.1897274	0.019421	-0.2097575	0.3328065	-1.1211	-1.46346	-0.918636
3808893	-0.050107724	-0.0783904	0.0493888	0.1003374	-0.0650181	0.1986185	0.386563	0.4117575	0.251207	0.379612	-1.53941	-2.548785	-0.645088
3808895	0.012452	0.0396339	0.4452735	0.3080875	0.104842396	0.506908	0.611934	0.478368	0.2969985	0.289638	0.1895935	-0.272819	0.146758
3789180	0.1535755	-0.162315	-0.0627405	0.0635337	0.4759885	0.0866629	0.03561275	0.1015159	0.4183365	0.365385	-0.2320115	-0.12659685	-0.264878
3808898	-0.9479155	0.01626355	-0.5673055	-0.422269	0.15540765	0.740076	0.4694455	0.800766	1.049149	0.8414505	-2.79763	-3.013195	-2.61144

Supplementary Table VI, Expression of 38 Aceview human TCF4 transcript variants and its Ensembl and UCSC isoforms.

Aceview mRNA variants	Ensembl isoform *	UCSC isoform *	N° exones	N° clones	From tissue (no strict specificity is implied)	5' completeness evidence	3' completeness evidence	minimal set of supporting clones	coordinates on gene
aApr07	ENST00000356073	uc002lfz.2	21	5	testis(4)	no evidence	validated polyA	BC031056 DB451911	1 to 408091
bApr07			20	1		no evidence	3' stop	BC125084	47985 to 408144
cApr07	ENST00000356073	uc002lfz.2	20	158	mandible, pooled (81), brain (26), eye (10), germinal center B cell (9) and 49 other tissues	capped	validated polyA	CR612521 AK095066 M74719 AK122765 AA292599 BE857360 F03966	47400 to 413697
dApr07	ENST00000356073	uc002lfz.2	18	24	brain (4), B cells (ramos cell line) (2), frontal lobe (2), lung (2) and 20 other tissues	capped	validated polyA	AK096862	125260 to 408339
eApr07			18	4	embryonic stem cells, embryoid bodiesderived from H1, H7 and H9 cells (2), B cells (ramos cell line) (1), thymus (1)	capped	validated polyA	CR624281 DB106801	47412 to 408038
fApr07	ENST00000398339 ENST00000354452	uc002lga.2 uc010dph.1	16	7	brain (3), neuroblastoma, cell line (2), dorsal root ganglia (1), embryonic stem cells, embryoid bodiesderived from H1, H7 and H9 cells (1) and 5 other tissues	5' stop	3' stop	AK095041	232033 to 408173
gApr07			9	2	lymph (4)	5' stop	3' stop	AW575235 AW402684	314232 to 408187
hApr07	ENST00000356073	uc002lfz.2	13	61	bone marrow (12), normal nasopharynx (12), brain (9), lung (4) and 22 other tissues	5' stop	validated polyA	M74718 BP350339	313591 to 408345
iApr07			13	2	brain (2), neuroblastoma (2), neuroblastoma cells (2), whole embryo, mainly head (2) and 2 other tissues		3' stop	AL518902 AL518903	333407 to 408338
jApr07			13	4	dorsal root ganglia (1), neuroblastoma (1)	5' stop	3' stop	CR614823	333407 to 408150

kApr07	ENST00000398339 ENST00000354452	uc002lga.2 uc010dph.1	14	94	brain (14), whole brain (7), kidney (4), lung (4) and 53 other tissues	5' stop	validated polyA	AB209741	234177 to 408913
lApr07			6	2	bone (1), ovary (pool of 3) (1), subchondral bone (1)	no evidence	3' stop	BM805150 BU615862	378533 to 408182
mApr07			12	7	brain (2), heart (2), liver and spleen (2)	capped	3' stop	BX401697 DA784720 W69216 W69096	47837 to 374938
nApr07			9	2	chondrosarcoma grade II (1), left pelvis (1)	no evidence	validated polyA	AU279553 BQ574248	89094 to 376400
oApr07			12	9	nervous normal (3), testis (2), amygdala (1), pooled human melanocyte, fetal heart, andpregnant uterus (1) and 2 other tissues	5' stop	3' stop	CR933675 DB443719	56 to 371422
pApr0			8	1	leiomyosarcoma (1), uterus (1)	no evidence	Partial	BQ435271	314288 to 381497
qApr07			9	2	coronary artery (1), leiomyosarcoma cell line (1), uterus (1)	no evidence	Partial	BG337255 BP206418	47829 to 360321
rApr07			7	20	brain (11), hippocampus (4), amygdala (2), ovary (1) and 4 other tissues	capped	Partial	DB727900	232265 to 366170
sApr07			7	2	brain (2)	no evidence	Partial	BP230971 BP230319	213731 to 366086
tApr07			7	1	muscle (1), rhabdomyosarcoma (1)	5' stop	3' stop	BF309122	314218 to 381387
uApr07			9	18	brain (10), dorsal root ganglia (1), embryonic stem cells, cell lines H1, H7, andh9 (1), thyroid (1) and 2 other tissues	capped	validated polyA	AA626251 BP265842	46218 to 304639
vaApr07			7	1	testis (1)	5' stop	Partial	BG773334	155790 to 356460
vbApr07			6	2	coronary artery (1), eye (1), rpe/choroid (1)	5' stop	Partial	BP234628	234693 to 366197
vcApr07			8	2	embryonic stem cell, retinoic acid andmitogen-treated hes cell line H7 (1), lung (1)	5' stop	Partial	CN404348	49909 to 356415
vdApr0			5	1	testis (1)	capped	Partial	DB040143	81 to 171913
veApr07			7	2	cerebrum (1), coronary artery (1)	5' stop	Partial	BP211580	47460 to 285135
vfApr07			7	1	cerebrum (1)	5' stop	Partial	BP214032	314733 to 378655
vgApr07			4	1		no evidence	3' stop	AA394023	403453 to 408025
vhApr07			5	1	brain (1)	5' stop	3' stop	BP230382	284042 to 360379
viApr07			6	3	brain (3)	5' stop	Partial	CB155670	232329 to 366088
vjApr07			5	3	brain (3)	5' stop	3' stop	BP348534	232368 to 360368

vkApr07			5	2	embryo, 8 weeks (1), normal nasopharynx (1)	5' stop	Partial	CD691430	213760 to 356477
vlApr07			7	1	testis (1)	capped	Partial	DB055535	137 to 285034
vmApr07			5	1	stomach (1)	5' stop	Partial	BM755137	50058 to 232551
vnApr07-unspliced			1	4	lung (1), thymus, pooled (1)	no evidence	No evidence	BG546663	46252 to 47319
voApr07			7	1	stomach (1)	no evidence	No evidence	BM787123	50062 to 285061
vpApr07-unspliced			1	1	whole embryo, mainly head (1)	no evidence	No evidence	DB354708	374387 to 374937
vqApr07-unspliced			1	1	head neck (1)	no evidence	No evidence	BG993333	304940 to 305222

* Ensembl and UCSC isoform that share the same supporting clones with AceView mRNA variants.

Supplementary Table VII, Enrichment analysis for the genes that are coexpresed with TCF4 in human cortex.

Pathway	p-value	Network objects	Genes	tissue
Development_EDG1 signaling pathway	0.0006286	3/21	G-protein alpha-o; GSK3 beta; PI3K reg class IA (p85-alpha)	Cortex
Development_Alpha-2 adrenergic receptor activation of MAPK1/3	0.001489	3/28	Alpha-2A adrenergic receptor; G-protein alpha-o; PI3K reg class IA (p85-alpha)	Cortex
G-protein signaling_Rac3 regulation pathway	0.002907	2/10	ABLIM1; Neurabin-1	Cortex
Cell cycle_Nucleocytoplasmic transport of CDK/Cyclins	0.003535	2/11	GSK3 beta; Karyopherin beta 1	Cortex
Cytoskeleton remodeling_ACM3 and ACM4 in keratinocyte migration	0.004962	2/14	ACM3; G-protein alpha-o	Cortex
Development_WNT signaling pathway. Part 2	0.005864	3/45	GSK3 beta; ITF2; PP2C alpha	Cortex
Cell adhesion_Synaptic contact	0,002909	6/127	CASK; GABA-A receptor alpha-1 subunit; GABA-A receptor alpha-2 subunit; GluR3; Neurexin 1-alpha; Neurexin 1-beta	Cortex
Reproduction_Progesterone signaling	0,006381	6/156	DYNC1I1; GABA-A receptor alpha-1 subunit; GABA-A receptor alpha-2 subunit; GSK3 beta; NCOA3 (pCIP/SRC3); PI3K reg class IA (p85-alpha)	Cortex

Supplementary Table VIII UCSC Tables used to define functional regions within TCF4.

Name of the track	Description
Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)	This track shows multiple alignments of 28 vertebrate species and two measures of evolutionary conservation -- conservation across all 28 species and an alternative measurement restricted to the placental mammal subset (17 species plus human) of the alignment. These two measurements produce the same results in regions where only mammals appear in the alignment. For other regions, the non-mammalian species can either boost the scores (if conserved) or decrease them (if non-conserved). The mammalian conservation helps to identify sequences that are under different evolutionary pressures in mammalian and non-mammalian vertebrates.
Human mRNAs from GenBank	The mRNA track shows alignments between human mRNAs in GenBank and the genome.
Human ESTs Including Unspliced	This track shows alignments between human expressed sequence tags (ESTs) in GenBank and the genome.
EvoFold Predictions of RNA Secondary Structure	This track shows RNA secondary structure predictions made with the EvoFold program, a comparative method that exploits the evolutionary signal of genomic multiple-sequence alignments for identifying conserved functional RNA structures.
Non-coding RNA Genes (dark) and Pseudogenes (light)	This track shows the location of non-protein coding RNA genes and pseudogenes.
FirstEF: First-Exon and Promoter Prediction	This track shows predictions from the FirstEF (First Exon Finder) program. Three types of predictions are displayed: exon, promoter and CpG window.
Vista HMR-Conserved Non-coding Human Enhancers from LBNL	The VISTA Enhancer Browser identifies distant-acting transcriptional enhancers in the human genome by coupling the identification of evolutionary conserved non-coding sequences with a moderate throughput mouse transgenesis enhancer assay.
TargetScan miRNA Regulatory Sites	This track shows conserved mammalian microRNA regulatory target sites for conserved microRNA families in the 3' UTR regions of Refseq Genes, as predicted by TargetScanS.
ENCODE NHGRI Elnitski Bidirectional Promoters	Annotation of bidirectional promoters, using in silico approaches.
CpG Islands (Islands < 300 Bases are Light Green)	This track shows CpG islands regions.
Eur. Inst. Oncology/J. C. Venter Inst. Nuclease Accessible Sites	The track annotates the location of NAS in the genome of human CD34+ and CD34- cells in the form of tags, generated by NA-Seq and obtained by merging NAS within 600 bp.
Eponine Predicted Transcription Start Sites	The Eponine program provides a probabilistic method for detecting transcription start sites (TSS) in mammalian genomic sequence, with good specificity and excellent positional accuracy.
Regulatory elements from ORegAnno	This track displays literature-curated regulatory regions, transcription factor binding sites, and regulatory polymorphisms from ORegAnno (Open Regulatory Annotation).
SwitchGear Genomics Transcription Start Sites	This track describes the location of transcription start sites (TSS) throughout the human genome along with a confidence measure for each TSS based on experimental evidence.
FOX2 adaptor-trimmed CLIP-seq reads	The FOX2 CLIP-seq track shows adaptor-trimmed CLIP-seq reads that mapped uniquely to the repeat-masked human genome (hg17). The reads were converted to hg18 coordinates using the UCSC LiftOver tool. Reads on the forward strand are displayed in blue; those on the reverse strand are shown in red.
NKI Nuclear Lamina Associated Domains (LaminB1 DamID) Tracks	The LaminB1 track shows a high resolution map of the interaction sites of the entire genome with Lamin B1, (a key NL component) in human fibroblasts..
Uppsala University ChIP-chip Tracks	This super-track combines related tracks of genome-wide ChIP-chip data generated by the Wadelius lab at Uppsala University, Sweden. These tracks display localization of two transcription factors (USF1 and USF2) and acetylated histone H3 (H3ac) in a liver cell line (HepG2), as assayed by Affymetrix arrays tiled at 35 bp resolution. For each factor, the average probe level intensities as well as identified positive regions are presented.
ENCODE Transcription Factor Binding Sites by ChIP-seq from Yale/UC-Davis/Harvard	This track shows probable binding sites of the specified transcription factors (TFs) in the given cell types as determined by chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq).

ENCODE Open Chromatin, Duke/UNC/UT	These tracks display evidence of open chromatin in multiple cell types from the Duke/UNC/UT-Austin/EBI ENCODE group. Open chromatin was identified using two independent and complementary methods: DNaseI hypersensitivity (HS) and Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE), combined with chromatin immunoprecipitation (ChIP) for select regulatory factors. Each method was verified by two detection platforms: Illumina (formerly Solexa) sequencing by synthesis, and high-resolution 1% ENCODE tiled microarrays supplied by NimbleGen.
ENCODE Univ. Washington DNaseI Hypersensitivity by Digital DNaseI	This track shows DNaseI sensitivity measured genome-wide in different cell lines using the Digital DNaseI methodology, and DNaseI hypersensitive sites. For each experiment (cell type) this track shows DNaseI hypersensitive zones (*HotSpots*) and hypersensitive sites (*Peaks*) based on the sequencing tag density (*Signal*).
ENCODE Univ. Washington Digital DNase Genomic Footprinting	This track, produced as part of the ENCODE Project, contains deep sequencing DNase data that will be used to identify sites where regulatory factors bind to the genome (*footprints*).
UW Predicted Nucleosome Occupancy Tracks	This track contains predicted nucleosome occupancy scores produced by three different computational models. Each model can predict regions of high and low nucleosome occupancy, one model (MEC) excels at recognizing regions of low nucleosome occupancy, whereas the other two (A375 and Dennis) are better at recognizing regions of high nucleosome occupancy.
ENCODE SUNY Albany RNA Binding Proteins by RIP-chip	This track shows expression of target RNA binding proteins (RBPs) as measured by RNA-binding protein immunoprecipitation-microarray profiling (RIP-chip) using different RIP antibodies in multiple cell lines.
GIS ChIP-PET	This track shows binding sites for p53, STAT1, c-Myc, histone modifications H3K4me3 and H3K27me3, as determined by chromatin immunoprecipitation (ChIP) and paired-end di-tag (PET) sequencing.
ESPERR Regulatory Potential (7 Species)	This track displays regulatory potential (RP) scores computed from alignments of human, chimpanzee (panTro2), macaque (rheMac2), mouse (mm8), rat (rn4), dog (canFam2), and cow (bosTau2). RP scores compare frequencies of short alignment patterns between known regulatory elements and neutral DNA.
HMR Conserved Transcription Factor Binding Sites	This track contains the location and score of transcription factor binding sites conserved in the human/mouse/rat alignment.
Ludwig Institute/UC San Diego TAF1 Binding in Fibroblasts Tracks	This super-track combines related tracks of genome-wide ChIP-chip data generated by the Ludwig Institute/UC San Diego ENCODE group. These tracks show TAF1 binding in fibroblastoid (IMR90) cells as assayed by ChIP-chip using a NimbleGen microarray.

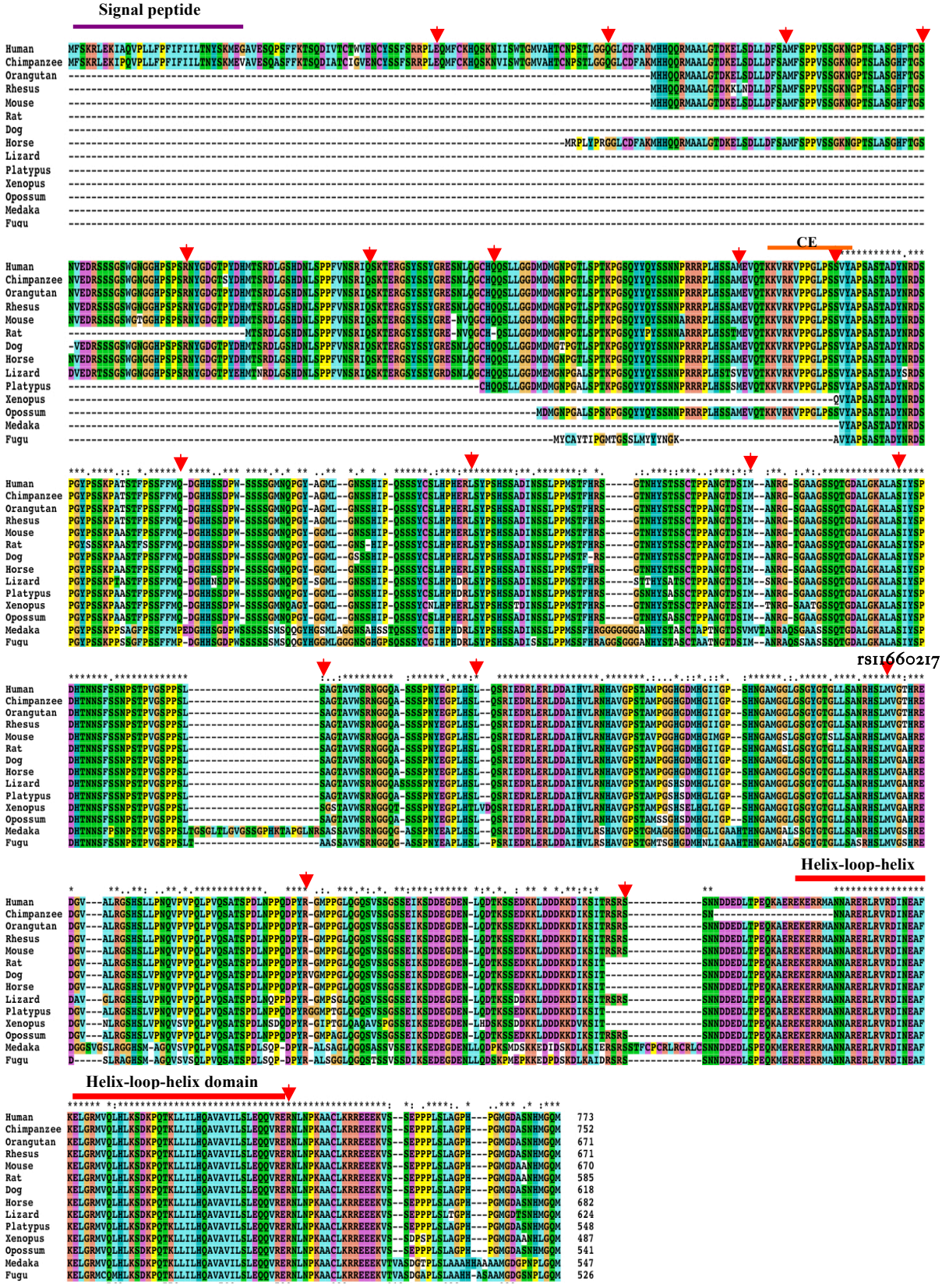
Supplementary Table IX SNP within exons in TCF4 gene region (chr. 18: 49019029 – 52414148, assembly hg18).

Variation ID	Chromosome name	Chromosome start	Chromosome end	Consequence to transcript
rs72496679	chr18	49036588	49095837	WITHIN_NON_CODING_GENE
rs34573287	chr18	49085952	49085952	NON_SYNONYMOUS_CODING
rs2271042	chr18	49086069	49086069	NON_SYNONYMOUS_CODING
rs35691189	chr18	49102465	49102465	NON_SYNONYMOUS_CODING
rs2278339	chr18	49120192	49120192	NON_SYNONYMOUS_CODING
rs3764495	chr18	49166492	49166492	SYNONYMOUS_CODING
rs984274	chr18	49190932	49190932	NON_SYNONYMOUS_CODING
rs34395300	chr18	49190991	49190991	SYNONYMOUS_CODING
rs35891220	chr18	49230895	49230895	SYNONYMOUS_CODING
rs34075746	chr18	49239605	49239605	SYNONYMOUS_CODING
rs2270950	chr18	49239778	49239778	NON_SYNONYMOUS_CODING
rs71679147	chr18	49275215	49285157	SYNONYMOUS_CODING
rs72918275	chr18	49279741	49279741	SYNONYMOUS_CODING
rs9957887	chr18	49922777	49922777	MISSENSE
rs61753467	chr18	49944955	49944955	SYNONYMOUS_CODING
rs10584411	chr18	50049957	50049957	NON_SYNONYMOUS_CODING
rs74180463	chr18	50049958	50049960	NON_SYNONYMOUS_CODING
rs3729509	chr18	50049964	50049966	FRAMESHIFT
rs60948705	chr18	50049964	50049964	FRAMESHIFT
rs3730669	chr18	50050000	50050000	SYNONYMOUS_CODING
rs3218778	chr18	50054337	50054337	NON_SYNONYMOUS_CODING
rs3218785	chr18	50058216	50058218	NON_SYNONYMOUS_CODING
rs3218788	chr18	50061120	50061120	SYNONYMOUS_CODING
rs3218784	chr18	50061257	50061257	NON_SYNONYMOUS_CODING
rs3218783	chr18	50063233	50063233	STOP_GAINED
rs35830201	chr18	50063361	50063361	SYNONYMOUS_CODING
rs11558769	chr18	50067700	50067700	NON_SYNONYMOUS_CODING
rs35602237	chr18	50067729	50067729	SYNONYMOUS_CODING
rs3730773	chr18	50067759	50067759	SYNONYMOUS_CODING
rs56152401	chr18	50072250	50072250	STOP_GAINED
rs3730823	chr18	50074032	50074032	NON_SYNONYMOUS_CODING
rs3218786	chr18	50074206	50074206	NON_SYNONYMOUS_CODING
rs55920515	chr18	50074207	50074207	NON_SYNONYMOUS_CODING
rs3218787	chr18	50074289	50074289	NON_SYNONYMOUS_CODING
rs8305	chr18	50074802	50074802	NON_SYNONYMOUS_CODING
rs2917782	chr18	50109718	50109718	NON_SYNONYMOUS_CODING
rs73958758	chr18	50112153	50112153	STOP_GAINED
rs17470798	chr18	50112194	50112194	SYNONYMOUS_CODING
rs17292725	chr18	50134886	50134886	STOP_GAINED
rs3819129	chr18	50141022	50141022	NON_SYNONYMOUS_CODING
rs1657904	chr18	50141120	50141120	SYNONYMOUS_CODING
rs1657907	chr18	50142166	50142166	NON_SYNONYMOUS_CODING
rs16958096	chr18	50142313	50142313	NON_SYNONYMOUS_CODING,SPLICE_SITE
rs73958766	chr18	50146113	50146113	SYNONYMOUS_CODING
rs35428499	chr18	50409545	50409545	NON_SYNONYMOUS_CODING
rs35802037	chr18	50416121	50416121	SYNONYMOUS_CODING

rs9947055	chr18	50416305	50416305	NON_SYNONYMOUS_CODING
rs3764511	chr18	50697671	50697671	SYNONYMOUS_CODING
rs9966265	chr18	50702595	50702595	NON_SYNONYMOUS_CODING
rs73956736	chr18	50707543	50707543	NON_SYNONYMOUS_CODING
rs34751112	chr18	50737542	50737542	NON_SYNONYMOUS_CODING
rs1344011	chr18	50756185	50756185	SYNONYMOUS_CODING
rs17852674	chr18	50756185	50756186	FRAMESHIFT
rs4638673	chr18	50756260	50756260	SYNONYMOUS_CODING
rs4294894	chr18	50756265	50756265	NON_SYNONYMOUS_CODING
rs72928889	chr18	50756276	50756276	NON_SYNONYMOUS_CODING
rs61743412	chr18	50760942	50760942	SYNONYMOUS_CODING
rs8766	chr18	51046528	51046528	SYNONYMOUS_CODING
rs71368997	chr18	51047271	51047271	SYNONYMOUS_CODING
rs11660217	chr18	51072725	51072725	NON_SYNONYMOUS_CODING
rs9947593	chr18	51599392	51599392	WITHIN_NON_CODING_GENE
rs10502989	chr18	50195818	50195819	UNKNOWN
rs16958240	chr18	50196362	50196363	UNKNOWN
rs16958245	chr18	50196444	50196445	UNKNOWN

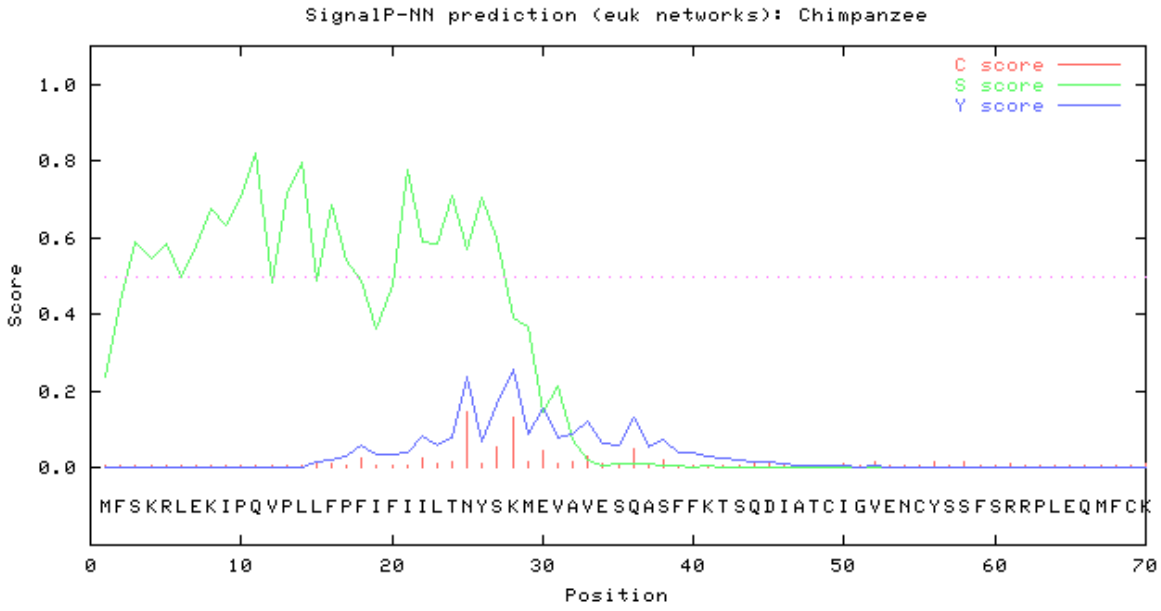
SUPPLEMENTARY FIGURES

Supplementary Figure 1 Alignment of the 14 TCF4 amino acid sequences. Intron positions are indicated by arrows. The positions of the introns were defined by the human sequence isoform ENST00000398339. The predicted signal peptide domain is labeled by a purple line, the CE element is labeled by an orange line and the Helix-loop-helix domain is labeled by a red thick line. One nonsynonymous SNP (rs11660217) in isoform is indicated.


Signal peptide
CE
rs11660217
Helix-loop-helix
Helix-loop-helix domain
Human
Chimpanzee
Orangutan
Rhesus
Mouse
Rat
Dog
Horse
Lizard
Platypus
Xenopus
Opossum
Medaka
Fugu

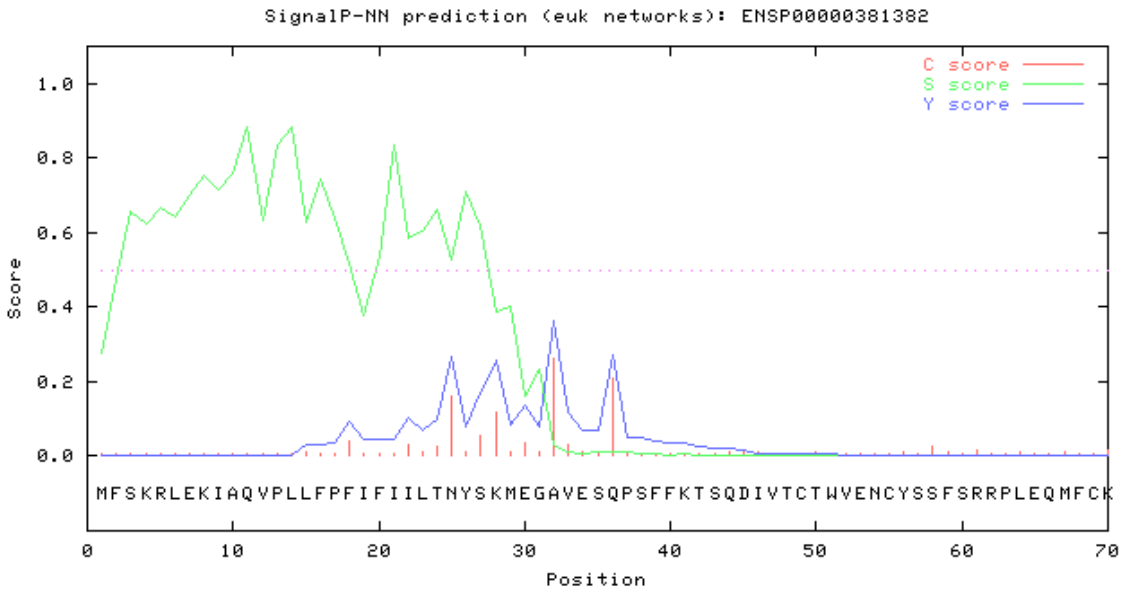
Supplementary Figure 2 SignalP-NN results for Human and Chimpanzee TCF4 sequences. The graphical output of SignalP-NN shows C-, S- and Y- score. The cleavage site is predicted to be at the position of maximal Y-score.

>Chimpanzee




```
>Chimpanzee            length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C     25       0.148   0.32   NO
  max. Y     28       0.257   0.33   NO
  max. S     11       0.823   0.87   NO
  mean S      1-27    0.588   0.48   YES
       D      1-27    0.423   0.43   NO
# Most likely cleavage site between pos. 27 and 28: NYS-KM
```

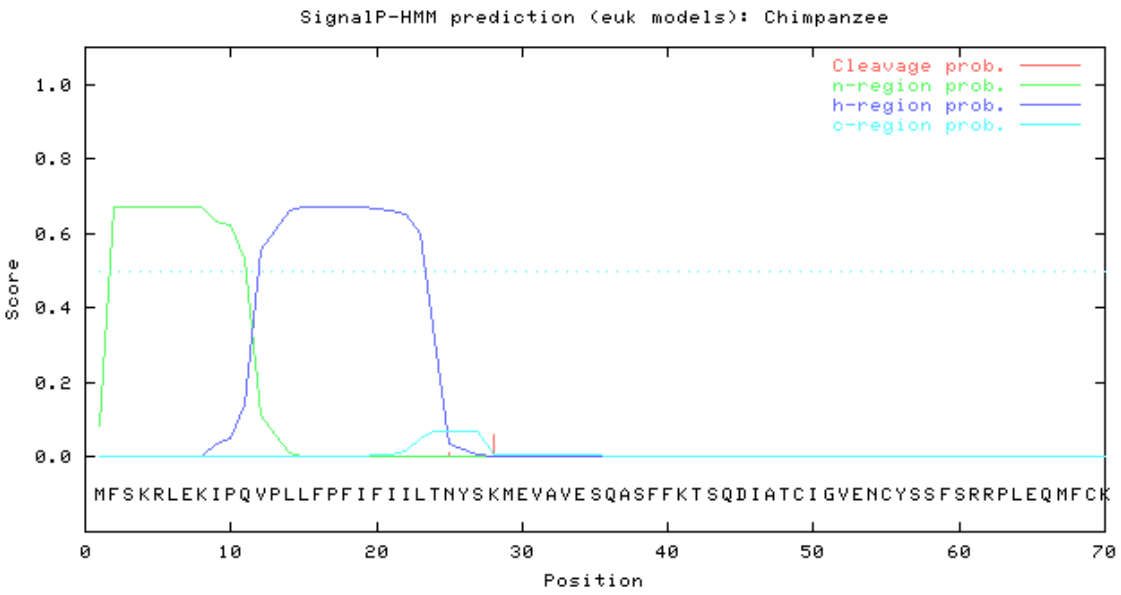
> Human (ENSP00000381382)




```
>ENSP00000381382       length = 70
# Measure  Position  Value  Cutoff  signal peptide?
  max. C     32       0.262   0.32   NO
  max. Y     32       0.363   0.33   YES
  max. S     11       0.883   0.87   YES
  mean S      1-31    0.602   0.48   YES
       D      1-31    0.482   0.43   YES
# Most likely cleavage site between pos. 31 and 32: MEG-AV
```

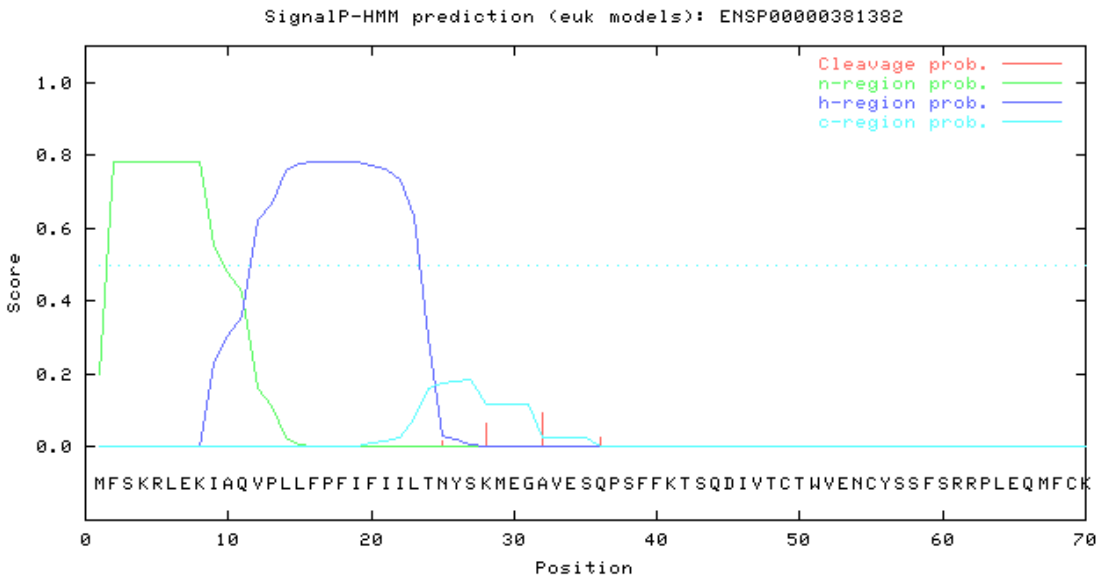
Supplementary Figure 3. SignalP-HMM result for Human and Chimpanzee TCF4 sequences. The graphical output of SignalP-NN shows the posterior probabilities for n-, h- and c- region and cleavage site.

>Chimpanzee




```
Prediction: Signal anchor
Signal peptide probability: 0.082
Signal anchor probability: 0.591
Max cleavage site probability: 0.061 between pos. 27 and 28
```

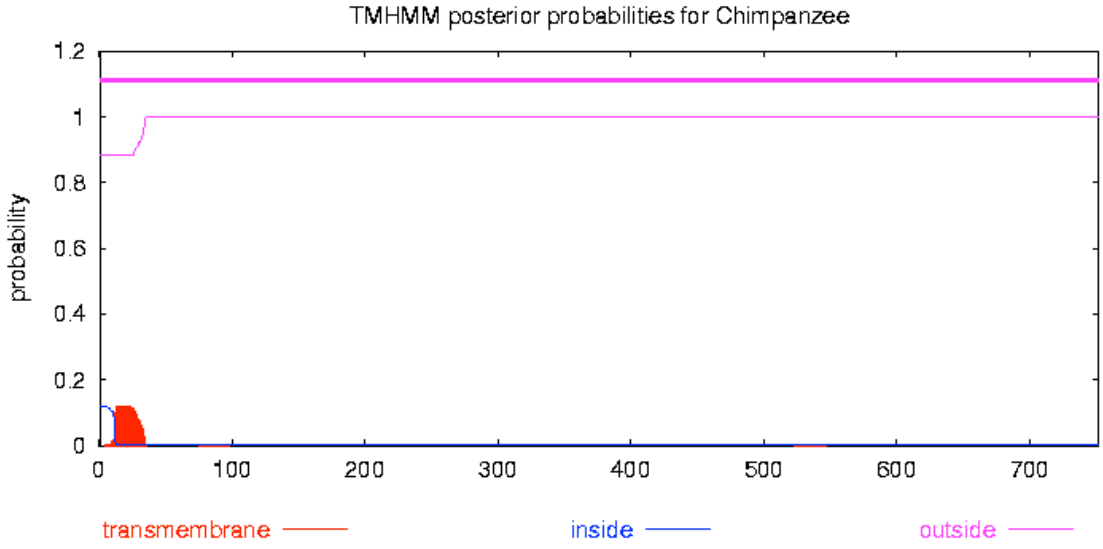
> Human (ENSP00000381382)



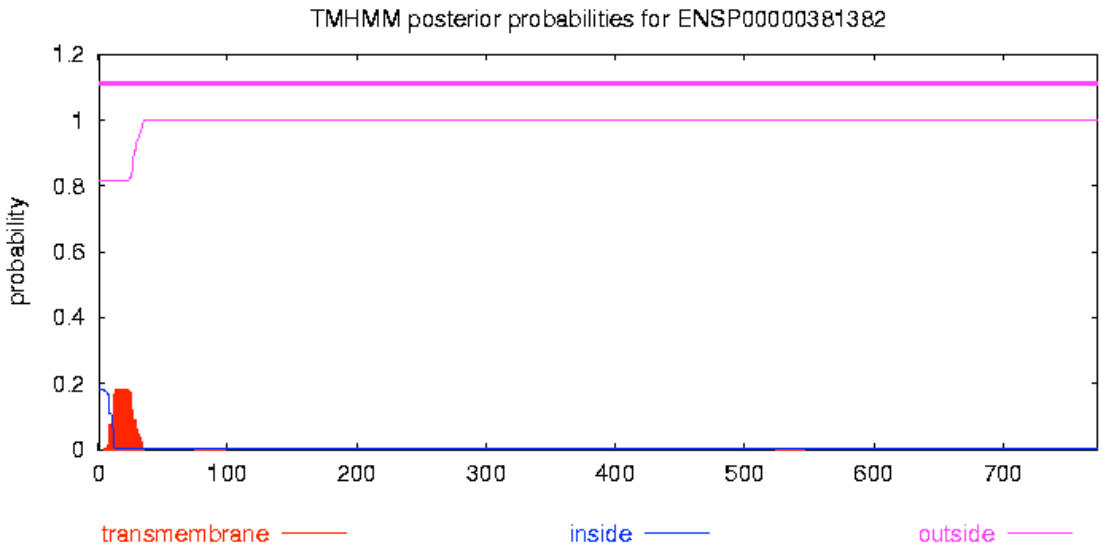
Prediction: Signal anchor
Signal peptide probability: 0.201
Signal anchor probability: 0.582
Max cleavage site probability: 0.092 between pos. 31 and 32

Supplementary Figure 4. The graphical output of TMHMM shows the probabilities for transmembrane, inside (i.e., cytoplasmatic), and outside (i.e., luminal or exterior) regions for Human and Chimpanzee TCF4 sequences.

```
# Human-773 Length: 773
# Human-773 Number of predicted TMHs:  0
# Human-773 Exp number of AAs in TMHs: 3.51425
# Human-773 Exp number, first 60 AAs:  3.46179
# Human-773 Total prob of N-in:        0.18192
Human-773       TMHMM2.0        outside      1   773
```

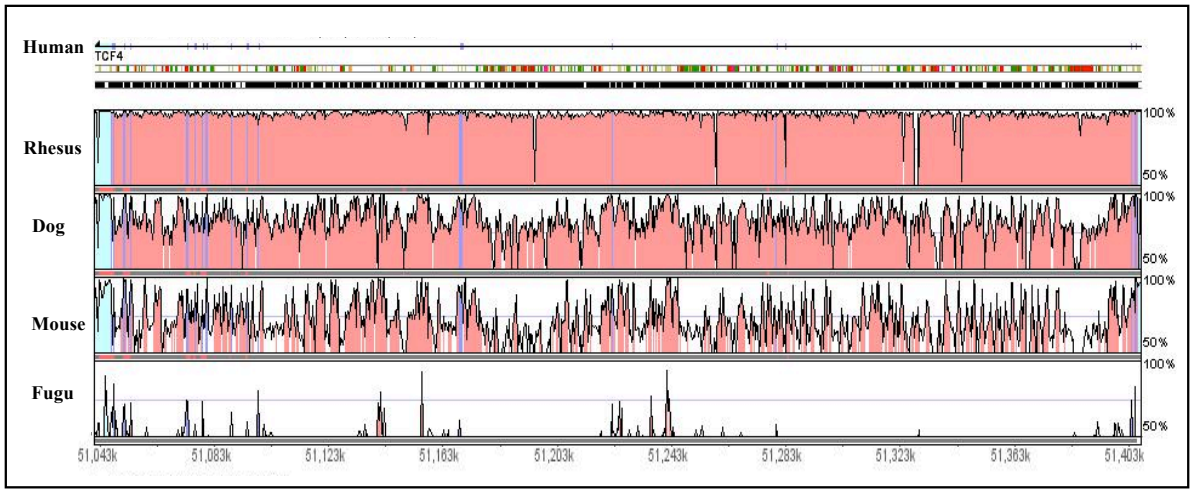



```
# Chimpanze-752 Length: 752
# Chimpanze-752 Number of predicted TMHs:  0
# Chimpanze-752 Exp number of AAs in TMHs: 2.39919
# Chimpanze-752 Exp number, first 60 AAs:  2.36479
# Chimpanze-752 Total prob of N-in:        0.11716
Chimpanze-752   TMHMM2.0        outside      1   752
```

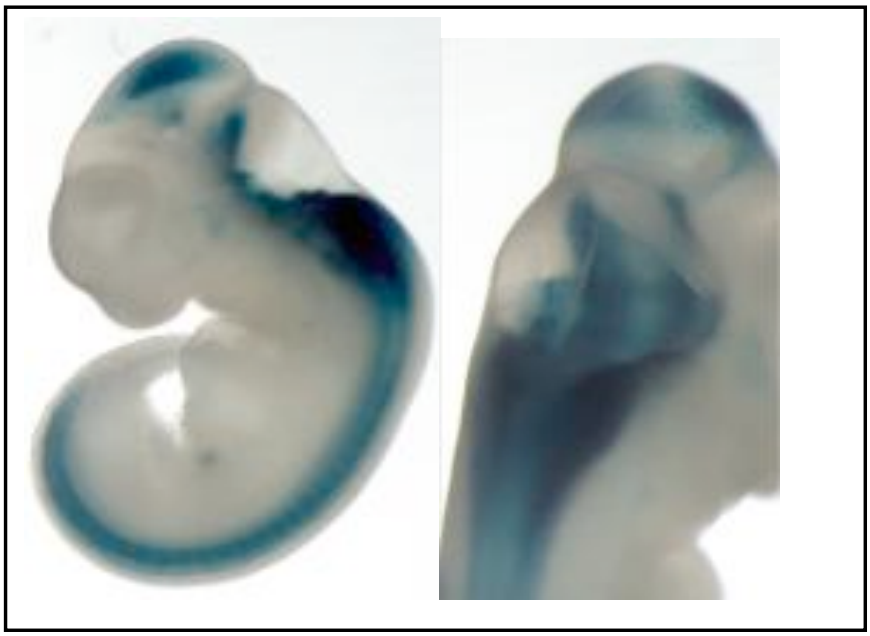


- Length: the length of the protein sequence.
- Number of predicted TMHs: The number of predicted transmembrane helices.
- Exp number of AAs in TMHs: The expected number of amino acids intransmembrane helices. If this number is larger than 18 it is very likely to be a transmembrane protein (OR have a signal peptide).
- Exp number, first 60 AAs: The expected number of amino acids in transmembrane helices in the first 60 amino acids of the protein. If this number more than a few, you should be warned that a predicted transmembrane helix in the N-term could be a signal peptide.
- Total prob of N-in: The total probability that the N-term is on the cytoplasmic side of the membrane.
- POSSIBLE N-term signal sequence: a warning that is produced when "Exp number, first 60 AAs" is larger than 10.

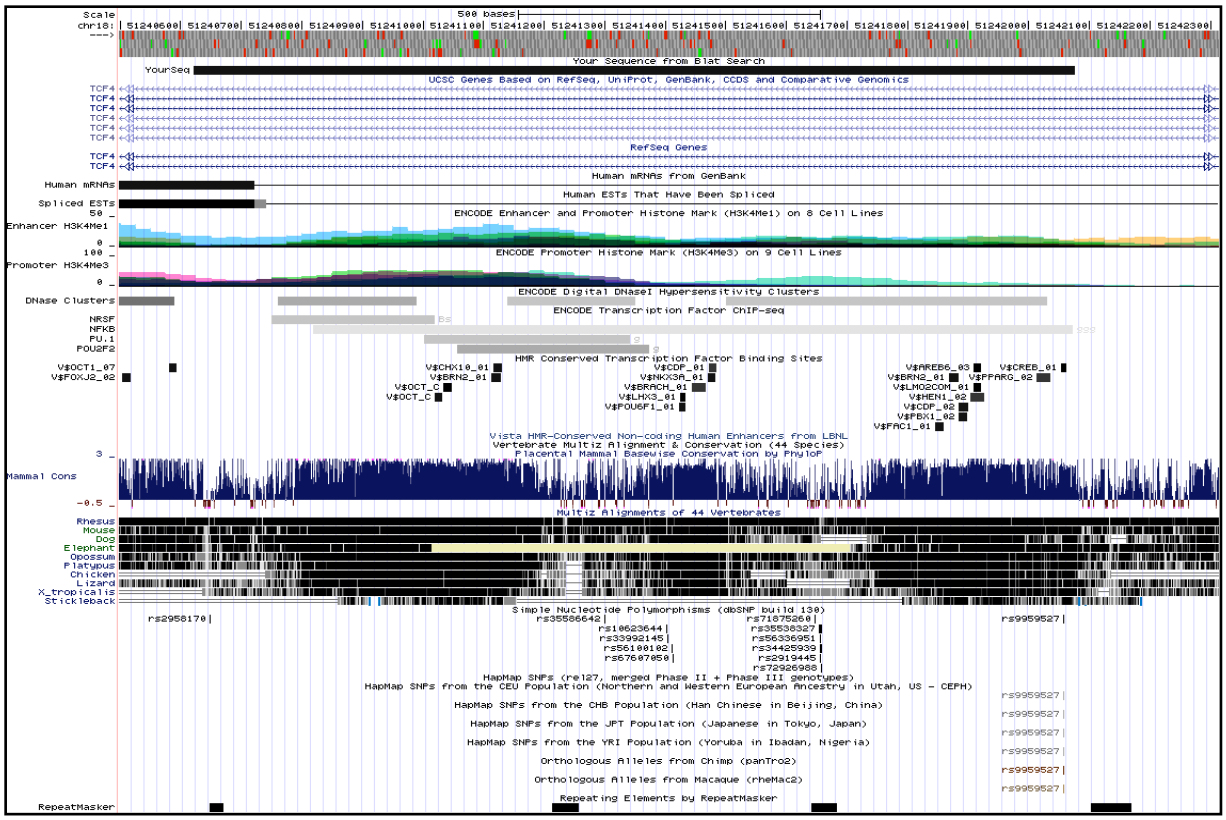
Supplementary Figure 5 VISTA plot displaying TCF4 gene structure and evolutionary conserved regions in a comparison of human, rhesus, dog, mouse and fugu DNA sequences. Multi-species conserved sequences (MCS, in red) is defined as an alignment at least 100-bp long and at least 70% identity. The plot used human gene and annotations as reference. The blue vertical lines denote an exon, and the red spaces between it denotes an intron.



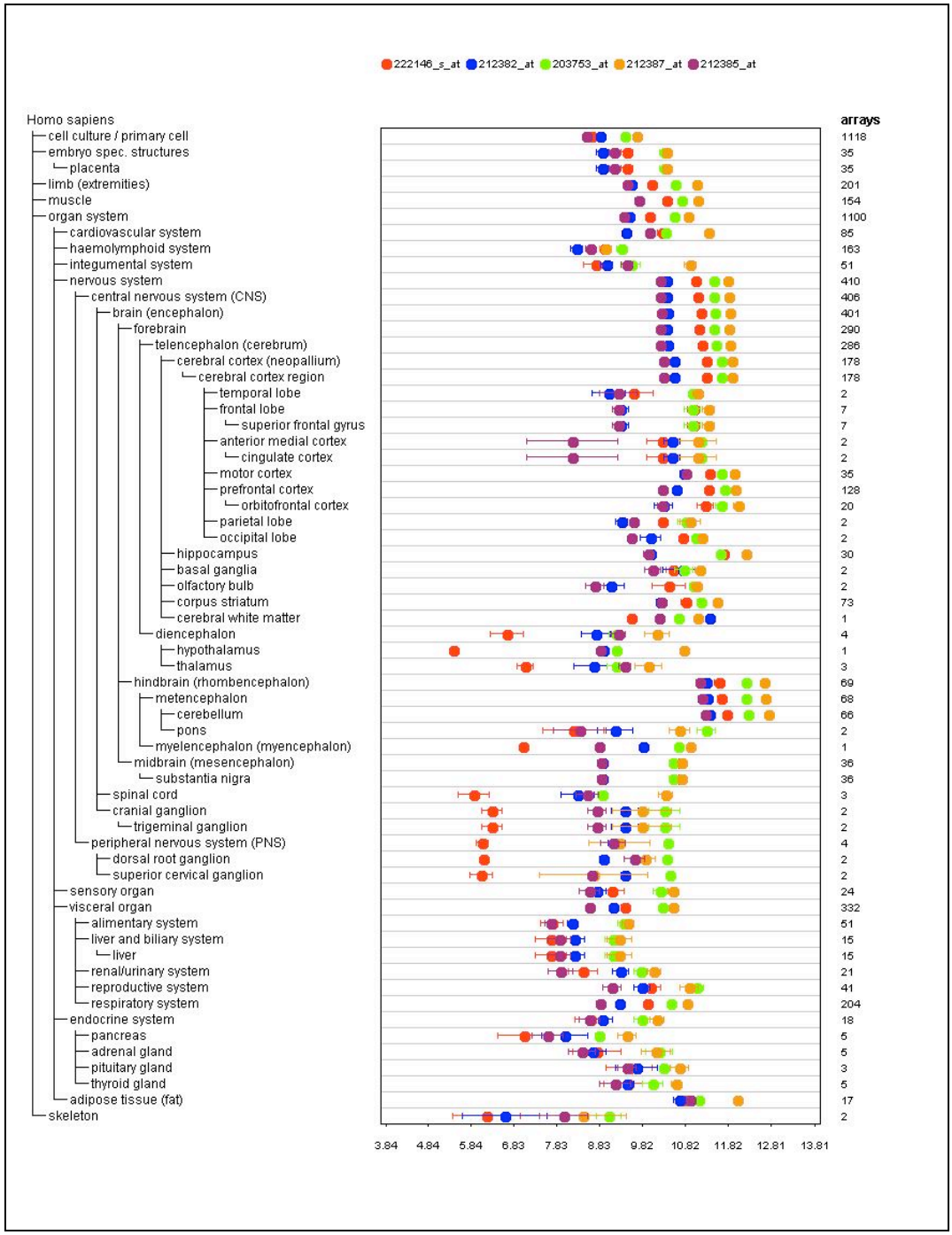
Suplementary Figure 6 Expression Pattern of Human element [hs376], Position: chr18:53,089,625-53,091,079, Source: Lawrence Berkeley National Laboratory, Flanking genes: TCF4 (intragenic)



Suplemetary Figure 7, enhancer region in UCSC, genetic variations within this regulatory element, and predicted transcription factor binding sites



Supplementary Figure 8 Expression of the five TCF4 probes in different tissues. Data were retrieved from Geneinvestigator.



Supplementary Figure 9 Expression of 212386_at TCF4 probe in different tissues. Data were retrieved from BioGPS.